# Journal of Experimental Psychology

165

Diary N. 73l(2)
Date ... 6.2
N
Research.

DAVID A. GRANT, Editor
*University of Wisconsin, Madison*

HAROLD W. HAKE, Associate Editor
*University of Illinois, Urbana*

LEONARD E. ROSS, Associate Editor
*University of Wisconsin, Madison*

LEO POSTMAN, Associate Editor
*University of California, Berkeley*

DELOS D. WICKENS, Associate Editor
*Ohio State University*

Consulting Editors

NANCY S. ANDERSON, *University of Maryland*
NORMAN H. ANDERSON, *University of California, San Diego*
WILLIAM H. BATCHELDER, *University of California, Irvine*
WILLIAM F. BATTIG, *University of Colorado*
LYLE E. BOURNE, JR., *University of Colorado*
GEORGE E. BRIGGS, *Ohio State University*
E. J. CAPALDI, *Purdue University*
VIRGIL R. CARLSON, *National Institute of Mental Health*
HERBERT H. CLARK, *Stanford University*
ANDREW CRIDER, *Williams College*
HELEN B. DALY, *State University of New York, College at Oswego*
ROGER L. DOMINOWSKI, *University of Illinois at Chicago Circle*
SHELDON M. EBENHOLTZ, *University of Wisconsin, Madison*
WARD EDWARDS, *University of Michigan*
HOWARD EGETH, *Johns Hopkins University*
BRUCE R. EKSTRAND, *University of Colorado*
HENRY C. ELLIS, *University of New Mexico*
JAMES R. ERICKSON, *Ohio State University*
MURRAY GLANZER, *New York University*
JUDITH GOGGIN, *University of California, Berkeley*
FRANCES K. GRAHAM, *University of Wisconsin*
WILLIAM W. GRINGS, *University of Southern California*
ROBERT C. HAYGOOD, *Arizona State University*
ELIOT HEARST, *Indiana University*
JULIAN HOCHBERG, *Columbia University*
LEONARD M. HOROWITZ, *Stanford University*
WILLIAM C. HOWELL, *Rice University*
STEWART H. HULSE, *Johns Hopkins University*
GEOFFREY KEPPEL, *University of California, Berkeley*
GREGORY A. KIMBLE, *University of Colorado*
HERBERT D. KIMMEL, *University of South Florida*
PATRICK R. LAUGHLIN, *University of Illinois*
HERSCHEL W. LEIBOWITZ, *Pennsylvania State University*
FRANK A. LOGAN, *University of New Mexico*
IRVING MALTZMAN, *University of California, Los Angeles*
DONALD R. MEYER, *Ohio State University*
BENNET B. MURDOCK, JR., *University of Toronto*
RAYMOND S. NICKERSON, *Bolt, Beranek, and Newman, Inc., Cambridge, Massachusetts*
MERRILL E. NOBLE, *Pennsylvania State University*
J. BRUCE OVERMIER, *University of Minnesota*
MICHAEL I. POSNER, *University of Oregon*
WILLIAM F. PROKASY, *University of Utah*
L. STARLING REID, *University of Virginia*
DONALD A. RILEY, *University of California, Berkeley*
WILLARD N. RUNQUIST, *University of Alberta*
RUDOLPH SCHULZ, *University of Iowa*
BRYAN E. SHEPP, *Brown University*
J. RICHARD SIMON, *University of Iowa*
PAUL SLOVIC, *Oregon Research Institute*
JANET T. SPENCE, *University of Texas, Austin*
HERBERT H. STENSON, *University of Illinois at Chicago Circle*
DAVID R. THOMAS, *University of Colorado*
MILTON A. TRAPOLD, *University of Minnesota*
ENDEL TULVING, *Yale University*
WILLIAM P. WALLACE, *University of Nevada*
WAYNE A. WICKELGREN, *University of Oregon*
GORDON WOOD, *Michigan State University*

---

HAROLD P. VAN COTT
Managing Editor

RITA JASKOWITZ
BETH FORRESTER
SUSAN REHR
Technical Editors

HENRY S. SALMON
Advertising and Business Manager

ANNE REDMAN
Subscription Manager

Bureau of Edn. & Psyl.
S. C. E. R. T.
Date 6-3-81
Acc. J 707

Volume 93, 1972

$10.00

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Second-class postage paid at Lancaster, Pa.

# CONTENTS OF VOLUME 93

Aiken, L. S., Ruskin, B., and Santa, J. L. Nonreinforced Trials in Concept Identification: Presolution Statistics and Local Consistency .... 100

Andreone, C. F. See Newman, F. L.

Arnold, P. G., and Bower, G. H. Perceptual Conditions Affecting Ease of Association .... 176

Barber, P. J., and Folkard, S. Reaction Time under Stimulus Uncertainty with Response Certainty .... 138

Barton, R. K. See Young, R. K.

Biederman, I. Human Performance in Contingent Information-Processing Tasks .... 219

Birnbaum, I. M. General and Specific Components of Retroactive Inhibition in the A-B, A-C Paradigm .... 188

Birnbaum, M. H. Morality Judgments: Test of an Averaging Model .... 35

Bower, G. H. See Arnold, P. G.

Briggs, G. E. See Swanson, J. M.

Brogden, W. J. See Fingeret, A. L.

Brotsky, S. J., and Litwin, L. M. Semantic Generalization: Replication of Cramer's Associative Gradient .... 430

Bruder, G. and Segal, E. Effects of Temporal and Spatial Organization of Lists on Clustering .... 151

Bugelski, B. R., and Lattanzio, S. A Culture-Free Learning Task .... 354

Carpenter, D. L. See Hershberger, W.

Chapman, C. R. and Feather, B. W. Modification of Perception by Classical Conditioning Procedures .... 338

Check, R. See Fox, R.

Checkosky, S. F. See Dumas, J.

Cole, L. E. See Kanak, N. J.

Daniel, T. C., and Ellis, H. C. Stimulus Codability and Long-Term Recognition Memory for Visual Form .... 83

Dooley, J. R. See Levin, I. P.

Dooling, D. J. Some Context Effects in the Speeded Comprehension of Sentences .... 56

Davidson, P. W. Haptic Judgments of Curvature by Blind and Sighted Humans .... 43

Deutsch, D. Effect of Repetition of Standard and Comparison Tones on Recognition Memory for Pitch .... 156

Disch, K. See Lappin, J. S.

Di Vesta, F. J. See Ingersoll, G. M.

Dobbs, A. R. Effect of Retention Interval, RI, and PI on Mediating Associations .... 417

Dong, T. Cued Partial Recall of Categorized Words .... 123

Drachman, D. A., and Leavitt, J. Memory Impairment in the Aged: Storage versus Retrieval Deficit .... 302

Dulberg, C. S. See Levin, I. P.

Dumas, J., Gross, E., and Checkosky, S. F. Effects of Attribute Probability in a Memory Search Task .... 327

Eckert, E. See Kanak, N. J.

Ellis, H. C. See Daniel, T. C.

Farris, J. C. See Geller, E. S.

Feather, B. W. See Chapman, C. R.

Fingeret, A. L., and Brogden, W. J. Item Arrangement Effects on Transfer and Serial Position Errors in Part–Whole Learning of Different Materials .... 249

Fisher, R. P. See Shulman, H. G.

Folkard, S. See Barber, P. J.

Fox, R., and Check, R. Independence between Binocular Rivalry Suppression Duration and Magnitude of Suppression .... 283

Fritzen, J. Memory Search and the Recall of Digit Series .... 410

Frost, N. See Umilta, C.

Gamboni, W. R., Gaustad, G. R., and Wilson, B. E. Verbal Discrimination Learning as a Function of Percentage Occurrence of Reinforcing Information Varying Presentation Rates .... 256

Gardner, G. T. See Shiffrin, R. M.

Gaustad, G. R. See Gamboni, W. R.

Geller, E. S. See Whitman, C. P. (a)

Geller, E. S. See Whitman, C. P. (b)

Geller, E. S. See Whitman, C. P. (c)

Geller, E. S., Whitman, C. P., and Farris, J. C. Probability Discrimination Indicated by Stimulus Predictions and Reaction Speed: Effects of S–R Compatibility (a) .... 404

Geller, E. S., Whitman, C. P., and McGuire, G. J. Incentive Effects in the Discrimination of Two Simultaneously Occurring Probability Distributions (b) .... 392

Gross, E. See Dumas, J.

Guralnick, M. J. Observing Responses and Decision Processes in Vigilance .... 239

Hammond, D. R. See Knowles, B. A.

Hawkes, L. See Spencer, T. J.

Heider, E. R. Universals in Color Naming and Memory .... 10

Heim, S. V., and Scholnick, E. K. Some Hypotheses about Negative Instances in Single-Attribute Concept Attainment .... 130

Hershberger, W. and Carpenter, D. L. Veridical Rotation in Depth in Unidimensional Polar Projections Devoid of Three Motion-Parallax Cues .... 213

Hinrichs, J. V. See Levin, I. P.

Hyman, R. See Umilta, C.

Ingersoll, G. M., and Di Vesta, F. J. Effects of Modality Preferences on Performance on a Bisensory Missing-Units Task .......... 386
Jahnke, J. C. Effect of Partial Recall on the Ranschburg Phenomenon .......... 118
Janke, M. Effect of Stimulus and Instructional Variables in an Ambiguous Concept-Attainment Task .......... 21
Johnsen, A. M. See Swanson, J. M.
Kanak, N. J., Cole, L. E., and Eckert, E. Implicit Associative Responses in Verbal Discrimination Acquisition .......... 309
Keele, S. W. Attention Demands of Memory Retrieval .......... 245
Knowles, B. A., Hammond, D. R., Stewart, T. R., and Summers, D. A. Detection of Redundancy in Multiple Cue Probability Tasks .......... 425
Lappin, J. S., and Disch, K. The Latency Operating Characteristic: II. Effects of Visual Stimulus Intensity on Choice Reaction Time .......... 367
Lattanzio, S. See Bugelski, B. R.
Leavitt, J. See Drachman, D. A.
LeMay, E. H. Stimulus Generalization Variables in Anagram-Problem Solving .......... 349
Levin, I. P., Dulberg, C. S., Dooley, J. R. and Hinrichs, J. V. Sequential Dependencies in Single-Item and Multiple-Item Probability Learning .......... 262
Litwin, L. M. See Brotsky, S. J.
Lively, B. L. The Von Restorff Effect in Short-Term Memory .......... 361
Lively, B. L., and Sanford, B. J. The Use of Category Information in A Memory-Search Task .......... 379
Mattson, G. See Spencer, T. J.
McGuire, G. J. See Geller, E. S.
Mewhort, D. J. K. Scanning, Chunking, and the Familiarity Effect in Tachistoscopic Recognition .......... 69
Milone, F. See Trumbo, D.
Murch, G. M. Binocular Relationships in a Size and Color Orientation Specific Aftereffect .......... 30
Nelson, T. O., and Rothbart, R. Acoustic Savings for Items Forgotten from Long-Term Memory .......... 357
Newman, F. I., Andreone, C. F., Washburn, L., and Purtle, R. B. Multidimensional Stimulus Control: Effects of Training and/or Testing .......... 290
Noble, M. See Trumbo, D.
Ornstein, P. A., and Schwartz, M. Encoding and Immediate Serial Recall of Consonant Strings .......... 105
Paivio, A. See Rowe, E. J.
Powell, A. See Weist, R. M.
Purtle, R. B. See Newman, F. L.
Ranken, H. B. See Santa, J. L.
Raser, G. A. False Recognition as a Function of Encoding Dimension and Lag .......... 333
Reutener, D. B. Background, Symbolic, and Class Shift in Short-Term Verbal Memory .. 90
Rothbart, R. See Nelson, T. O.
Rowe, E. J., and Paivio, A. Effects of Noun Imagery, Pronunciation, Method of Presentation, and Intrapair Order of Items in Verbal Discrimination .......... 427
Ruskin, B. See Aiken, L. S.
Segal, E. See Bruder, G.
Sanford, B. J. See Lively, B. L.
Santa, J. L. See Aiken, L. S.
Santa, J. L., and Ranken, H. B. Effects of Verbal Coding on Recognition Memory ...... 268
Scholnick, E. K. See Heim, S. V.
Schwartz, M. See Ornstein, P. A.
Shanteau, J. Descriptive versus Normative Models of Sequential Inference Judgment .... 63
Shaughnessy, J. J. See Underwood, B. J.
Shiffrin, R. M., and Gardner, G. T. Visual Processing Capacity and Attentional Control 72
Shulman, H. G., and Fisher, R. P. Expected Value As a Determinant of the Distribution of Attention .......... 343
Silverstein, A. Acquired Pleasantness and Paired-Associate Learning in Mixed and Homogeneous Lists .......... 111
Simpson, W. E. Latency of Locating Lights and Sounds .......... 169
Snyder, C. See Umilta, C.
Snyder, M. See Umilta, C.
Spencer, T. J., Hawkes, L., and Mattson, G. Effect of A Forward Indication on Backward Masking .......... 297
Steinberg, D. D. Truth, Amphigory, and the Semantic Interpretation of Sentences ...... 217
Stewart, T. R. See Knowles, B. A.
Summers, D. A. See Knowles, B. A.
Swanson, J. M., Johnsen, A. M., and Briggs, G. E. Recoding in a Memory Search Task .. 1
Thomson, W. J. Effect of Number of Response Categories on Dimension Selection, Paired-Associate Learning, and Complete Learning in a Conjunctive Concept Identification Task .......... 95
Trumbo, D., Milone, F., Noble, M. Interpolated Activity and Response Mechanisms in Motor Short-Term Memory .......... 205
Umilta, C., Frost, N., and Hyman, R. Interhemispheric Effects on Choice Reaction Times to One-, Two-, and Three-Letter Displays .......... 198
Umilta, C., Snyder, C., and Snyder, M. Repetition Effect As a Function of Event Uncertainty, Response-Stimulus Interval, and Rank Order of the Event .......... 320
Underwood, B. J., Shaughnessy, J. J., and Zimmerman, J. List Length and Method of Presentation in Verbal Discrimination Learning with Further Evidence on Retroaction 181
Wallace, R. J. Spatial S–R Compatibility Effects Involving Kinesthetic Cues .......... 163
Washburn, L. See Newman, F. L.

Weist, R. M., and Powell, A. Blocked Presentation in Multitrial Free Recall .......... 398
Whitman, C. P. See Geller, E. S. (a)
Whitman, C. P. See Geller, E. S. (b)
Whitman, C. P., and Geller, E. S. Prediction Outcome and Choice Reaction Time: Stimulus versus Response Anticipation (a) .......... 193
Whitman, C. P., and Geller, E. S. Sequential Effects of Stimulus Probability and Prediction Outcome on Choice Reaction Time (b) .......... 373
Whitman, C. P., and Geller, E. S. Stimulus Anticipation in Choice Reaction Time with Variable S–R Mapping (c) .......... 433
Wilson, B. E. See Gamboni, W. R.
Wilson, M. Assimilation and Contrast Effects in Visual Discrimination by Rhesus Monkeys 279
Young, R. K., and Barton, R. K. Manipulated Retrievability in Free Recall .......... 143
Zimmerman, J. See Underwood, B. J.

Diary N. 738(2)
Date 42
N. 116
Bureau Edul Psy: Research.

Journal of
Experimental Psychology

VOL. 93, NO. 1 APRIL 1972

# RECODING IN A MEMORY SEARCH TASK[1]

JAMES M. SWANSON,[2] ARTHUR M. JOHNSEN, AND GEORGE E. BRIGGS[3]
*Ohio State University*

By using the M. I. Posner and S. W. Keele physical versus name identity conditions in S. Sternberg's stimulus-classification task, linear functions were obtained relating reaction time to positive set size when (*a*) the memorized positive set items and the test item were of the same format (physical identity) and (*b*) also when they differed in format (associational or name identity). The slopes of these functions differed significantly (90 msec/item vs. 121 msec/item). It is suggested that a recoding operation caused this slower rate of central processing: When the items for storage in active memory and the test item are presented in the same format, *S* can perform the comparison process directly; but when the formats differ, each memory item must be recoded into the format of the test item prior to the comparison process.

Posner and Keele (1967) have shown that reaction time in a same–different task is shorter when the two stimuli being compared match on the basis of physical identity than when they match on the basis of associational (or name) identity; e.g., *S*s can respond "same" faster when two capital "As" are presented as stimuli than when a capital A and a lowercase a are presented. To account for this, Posner and his coworkers (Posner, Boies, Eichelman, & Taylor, 1969; Posner & Mitchell, 1967) have formulated a theory which is based on the construct "nodes of processing," a term which refers to the basis on which a match is made in memory. A physical identity match can occur at the first level, Node 1, on the basis of the physical correspondence between the two stimuli presented; however, if the two stimuli are physically dissimilar, a match can still occur but now it is dependent upon a learned correspondence between the two stimuli. A match of the latter sort (an associational identity match) occurs at the second level, Node 2. In the physical identity condition, when two identical stimuli are presented, they are encoded and represented identically in memory, so the comparison of the two stimuli can be made directly. However, in the associational identity condition, when two physically dissimilar stimuli are presented for comparison, they are encoded and represented in different formats in memory, so at least one of the encoded

---

[1] This research was supported in part by the National Science Foundation through Grant GN-534.1 from the Office of Scientific Information Service to the Computer and Information Science Research Center, and in part by Grant 2 TO1 NH 08526 from the United States Department of Health, Education, and Welfare and Grant OEG-5-9-450276-0053 from the United States Office of Education. A version of this study was presented at the Midwestern Psychological Association, Cincinnati, May 1970.

[2] Now at the University of Texas at Austin, Austin, Texas 78712.

[3] Requests for reprints should be sent to George E. Briggs, Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

© 1972 by the American Psychological Association, Inc.

representations must be recoded for the two stimuli to be represented in the same format in memory. According to this interpretation, then, a recoding operation is the difference between the Node 1 and Node 2 match in the "nodes of processing" theory.

The Posner and Keele (1967) results may be interpreted also in terms of Sternberg's (1967) model of character recognition. Sternberg identifies two stages of processing: (*a*) stimulus encoding and (*b*) memory comparison. During Stage 1, information is extracted from the stimulus and an internal representation of that stimulus is formed. During Stage 2, the encoded version of the test stimulus is compared to memorial representations of alternative stimuli which could have been presented.

Given Sternberg's (1967) model and results of Posner and Keele (1967), one may ask what is recoded when an associational identity match is made. Is the input from Stage 1 (the encoded version of the test stimulus) recoded into the format of the memorized target stimulus, or is the memorized version of the target stimulus recoded into the format of the test stimulus? This question cannot be answered from data collected in the same–different paradigm. However, the Sternberg (1967) methodology does offer a way to specify the locus of the recoding operation: On each trial, a positive set is defined by serially presenting a given number ($M$) of stimuli to $S$ to be memorized. Following presentation of the positive (memorized) set, a single test stimulus occurs, and $S$ is required to classify the test stimulus by making a positive response if the test stimulus is a member of the positive set or a negative response otherwise.

Sternberg's (e.g., 1969a) data show that $M$ and reaction time (RT) are linearly related,

$$\mathrm{RT} = a + b(M).$$

On the basis of these and other data (see Sternberg, 1969b), he proposed that the decision process in the stimulus-classification task consists of the two operations mentioned before: *stimulus encoding* and *memory comparison*. Since only one test stimulus is presented on each trial, the stimulus encoding operation occurs only once regardless of the size of the positive set. Thus, the time taken to encode the test stimulus is constant across $M$ and so it is included in the intercept of the RT memory load equation. But, according to Sternberg's theory, the memory comparison operation consists of successively and exhaustively comparing the encoded test stimulus to each of the memorized positive set items. This operation occurs once for each item in the positive set, and thus the time for one comparison determines the slope constant $b$ of the RT memory load additivity statement. Since a same–different task requires one encoding and one comparison operation on each trial, the encoding and comparison times cannot be separated as in the Sternberg methodology.

Consider the difference between the physical and associational identity match conditions of Posner and Keele (1967) in light of Sternberg's methodology. If, in the associational identity condition, recoding takes place in Stage 1, then only *one* recoding operation is required (the single test stimulus is recoded), and the predicted difference in the RT functions for the physical and associational identity conditions would be with regard to the intercept constant $a$. On the other hand, if the recoding operation takes place in Stage 2, then one recoding operation for *each* memory item is required, and the predicted difference in the RT functions would be with regard to the slope constant $b$. If no recoding is necessary on the trials when the format of the test item differs from the format of the memory items, or if the same recoding operation is required on every trial, then the RT functions for the two conditions should not differ in either slope or intercept values.

## METHOD

*Subjects and procedure.*—Twenty-five students responded to advertisements in the campus paper and each served for 10 daily 30-min. sessions. Each was paid $1.25 per session for his service.

During the first week of the experiment, $S$ learned two-digit numerals as names for 16 eight-sided random forms. These forms were selected from those made available by Vanderplas, Sanderson, and Vanderplas (1965). On the first day of the experiment, the 16 random forms were separated into two lists of 8 forms each, and the forward associations in both lists were learned. On the learning trials, each form was back projected alone on a ground-glass screen before *S*. Two seconds after the form was presented, its assigned name appeared at its side for an additional 2 sec. A 2-sec. blank interval followed each S–R pair. Each list of 8 was learned to a criterion of two consecutive perfect trials.

During the second session, the two lists were combined and the single list of 16 form-name pairs was relearned using the same procedure and criterion described above. Also, a backward-association learning procedure was initiated during this session. This consisted of presenting each of the 16 names individually for 4 sec., and during this time, *S* attempted to point to the appropriate one of the 16 forms printed on a sheet which was located on a table before him. Then the appropriate form would appear on the screen beside the name for an additional 2 sec. to provide feedback concerning the accuracy of the response. During the remainder of the first week (three additional sessions), both the forward and backward associations were practiced and highly overlearned. In each of these sessions, the list of 16 form-name pairs was presented and practiced until both the forward and backward associations could be recited to a criterion of three consecutive errorless trials.

The experimental task was introduced on the first session of the second week. The Sternberg (1966) varied-set procedure was used to present, on a visual display, either one, two, or four random forms or one, two, or four, two-digit numerals (names) as positive set items to be maintained in short-term memory. The *S* then was tested with either a random form or a two-digit numeral. A positive response was required (*a*) if the test item matched a memory item on the basis of physical identity (both items were the same random form or the same two-digit numeral) *or* (*b*) if the test item matched a memory item on the basis of associational identity (one was a random form and the other was the name of that form). A negative response was required if the test item matched none of the memory items either on the basis of physical identity or associational identity.

During the second week of the experiment, each *S* performed in the stimulus-classification task for five sessions. The first session was considered to be practice, and the data to be reported below were collected during the remaining four sessions. Each session involved 96 trials. Each trial consisted of (*a*) a warning signal (the word "ready" appeared on the display), (*b*) the visual presentation of the memory items one at a time for 2 sec. each, followed by a .5-sec. blank interval, (*c*) another warning (the word "test" appeared on the display), and (*d*) visual presentation of the test stimulus for 2 sec. A timer started with the onset of the test stimulus and was stopped when *S* pressed one of two buttons to make his positive or negative response. The *E* recorded the accuracy of the response and the reaction time.

*Design.*—A repeated-measures design was used in which each *S* emitted 16 responses in each of the 24 conditions defined by the factorial combination of the four variables: (*a*) type of match (physical or associational identity), (*b*) format of the test stimulus (random form or two-digit numeral), (*c*) memory load (one, two, or four items in the positive set), and (*d*) response type (positive or negative). Across the 4 days of data collection (a total of 384 trials), counterbalancing of stimulus materials was employed such that in each of the 24 conditions, (*a*) each of the 16 random forms and the 16 two-digit numerals was presented once in each of the serial positions of the several positive sets, (*b*) each of the 16 forms and the 16 numerals was presented once as a test item, and (*c*) when a match occurred, the matched memory item was presented originally equally often in each serial position. The 384 trials resulting from the above design and counterbalancing procedure were divided into four blocks of 96 trials by randomly selecting 4 of the trials from each of the 24 conditions to assign to each of the four data-collection sessions. The 96 trials in each session were then randomized with the restriction that there were no runs greater than 4 of the same memory load condition or of the same response type.

It is important to note that the presentation format of the memory items and the presentation format of the test item were independent; therefore, on each trial of the experiment, *S* was uncertain as to what would be the format of the test item after the memory items had been presented.

*Apparatus.*—Each of the 16 random forms and two-digit numerals used as stimuli were photographed separately on 35-mm. film and mounted in 2 × 2 slides. Two Kodak RA-950 random access projectors were used alternately to present the slides. The visual angle subtended by the stimuli was 5°. A Hewlett-Packard 52236 timer was used to record the time elapsing between the onset of a test stimulus and *S*'s response. The response was made on a panel containing two buttons on which *S* rested his right and left index fingers. The two buttons were marked YES and NO to provide a positive (match) and a negative (no-match) response, respectively. The assignment of buttons to right and left fingers was alternated across (not within) the 25 *S*s. The *S* wore headphones and a boom microphone by which he could communicate with *E*. When not in communication with *E*, *S* heard white noise at 75 db. This noise served to mask extraneous sounds from the immediate laboratory area.

## RESULTS

A mean reaction time (RT) was calculated for each *S* on the basis of his correct responses in each of the 24 experimental conditions described above, and these scores were used as the basic data in the analyses of variance. To be consistent with Posner and Keele's (1967) treatment of the data, positive and negative responses were analyzed separately. The group

TABLE 1

AVERAGE REACTION TIME (IN MSEC.) AND PERCENTAGE OF ERRORS

| Type of match | RT | | |
|---|---|---|---|
| | $M = 1$ | $M = 2$ | $M = 4$ |
| Physical | | | |
| F-F match | 953.2<br>(.50) | 1096.7<br>(5.50) | 1241.5<br>(4.75) |
| No match | 974.0<br>(2.75) | 1162.3<br>(4.25) | 1368.5<br>(6.50) |
| N-N match | 819.6<br>(3.00) | 878.2<br>(3.25) | 1077.9<br>(7.75) |
| No match | 874.9<br>(2.25) | 971.2<br>(.75) | 1141.9<br>(3.25) |
| Associational | | | |
| F-N match | 778.4<br>(2.25) | 909.9<br>(4.00) | 1137.5<br>(7.50) |
| No match | 890.4<br>(1.50) | 1043.3<br>(3.50) | 1207.6<br>(4.50) |
| N-F match | 985.8<br>(3.00) | 1152.9<br>(4.00) | 1365.5<br>(11.00) |
| No match | 1022.3<br>(4.25) | 1197.3<br>(.50) | 1433.3<br>(7.75) |

Note.—The format of the positive set and the format of the test stimulus are indicated by $F$ = form and $N$ = numeral. Percentage of errors are given in parentheses. $M$ = positive set size.

average RTs and percentages of error for the 24 conditions are shown in Table 1.

In the analysis of the positive response data, the basic findings of Posner and Keele (1967) and Sternberg (1966) were replicated. First, physical identity matches are made faster than associational identity matches, with an overall difference of 44 msec. between the two levels of this variable: $F$ (1, 24) = 11.6, $p < .01$. Second, RT increases monotonically with $M$, $F$ (2, 48) = 208.2, $p < .01$. A strong linear relationship exists (linear memory load, $F$ (1, 48) = 396.8), and the linear component accounts for 95% of the variance of the memory load variable.[4]

The interaction of these two variables (Type of Match × Memory Load $F$ (2, 48) = 12.0, $p < .01$) indicates that the profiles of the RT memory load functions for the physical and the associational identity conditions are not parallel. The Format of Test Stimulus × Type of Match × Memory Load interaction did not attain significance, $F$ (2, 48) = 1.07, indicating that the difference in slope for the physical and associational identity conditions was the same when a random form was presented as a test stimulus as when a two-digit numeral was presented as the test stimulus. The best-fit linear equations are shown in Fig. 1.

Taken with the significance of the format of test stimulus main effect, these results indicate that the intercept of the RT memory load function is dependent upon format of the test stimulus (names are encoded more quickly than forms), and that the slope, regardless of the test stimulus format, is dependent upon whether a physical or an associational identity match was made (physical identity comparisons are made more quickly than associational identity comparisons).

It is interesting to note that the slope

---

[4] Apparently the *form* of the functional relationship between RT and memory load depends upon the format of the test stimulus. When a numeral was presented as a test stimulus, the increase in RT from $M = 2$ to $M = 4$ was 1.8 times as great as the increase from $M = 1$ to $M = 2$, indicating that a linear relationship provided a slightly better description ($r_{\text{lin}} = .9980$ vs. $r_{\text{log}} = .9882$) than a log relationship. However, when a random form was presented as a test stimulus, the increase in RT from $M = 2$ to $M = 4$ was only 1.2 times as great as in the increase from $M = 1$ to $M = 2$, indicating that a logarithmic relationship provided a slightly better description ($r_{\text{log}} = .9982$ vs. $r_{\text{lin}} = .9882$) of the data than a linear relationship. All things considered, either functional relationship provides a good fit for these data, but since only three levels of $M$ were used, it is difficult to establish which functional relationship is the best. But since the difference in profiles for the physical and associational identity conditions, regardless of the form of the functional relationship between RT and memory load, is the critical difference that localizes an effect in Stage 2 (see Smith, 1967, and Sternberg, 1969a), this ambiguity concerning a log versus a linear relationship is not critical. If a logarithmic functional relationship between RT and memory load is used when a random form was presented as a test stimulus, the memory functions presented below are obtained:

$$\text{RT} = .780 + .352\ (\text{Hc})\ \text{physical ID}$$
$$\text{RT} = .796 + .398\ (\text{Hc})\ \text{associational ID},$$

where Hc is a Shannon (log) statement of central processing uncertainty developed by Swanson and Briggs (1969). From the above statements of additivity in reaction time, the same conclusion reached on the basis of a linear relationship is also reached on the basis of a logarithmic relationship: the difference in the physical and associational identity conditions is in the slope of the RT memory load functions. Since either form of the functional relationship between RT and memory load provides a good fit, and since the basic conclusions drawn in the paper do not differ as shown above, linear functions will be used throughout this article.

for the two physical identity conditions are nearly equal: $b = .088$ sec/item for the name-name (NN) condition and $b = .093$ sec/item for the form-form (FF) condition. The slope for the FF condition is consistent with data from other experiments in which random forms were used as stimuli: Briggs (1969) obtained the RT memory load function $\mathrm{RT} = .602 + .092\ (M)$ using eight-sided random forms. The slope for two-digit numerals differs substantially from the 40 msec/item usually obtained by Sternberg when one-digit numerals are used. To investigate the obtained equality of slopes between random forms and numerals further, a study was conducted to see if (*a*) the slope for two-digit numerals could be replicated and (*b*) the slope for numerals was dependent upon the number of digits per numeral.[5] One-digit, two-digit, and three-digit numerals were used. The reaction time functions obtained are presented below:

$\mathrm{RT} = .392 + .043\ (M)$ one-digit numerals
$\mathrm{RT} = .360 + .083\ (M)$ two-digit numerals
$\mathrm{RT} = .380 + .114\ (M)$ three-digit numerals

These data suggest that the slope of the RT memory load functions when numerals are used as stimuli is dependent upon the number of digits per numeral—the slope seems to be about 40 msec. per digit for all three numeral sizes used. This suggests that the equality of slopes for random forms and names (numerals) was merely a fortunate occurrence due to the choice of two-digit numbers as names for the forms. However, the equality of slopes for forms and names is not the important aspect of the present set of data: The important relationship is the *difference in slopes* (regardless of the format of the test stimulus) for the physical and associational identity conditions, and this difference remains even in the light of data from the control experiment.

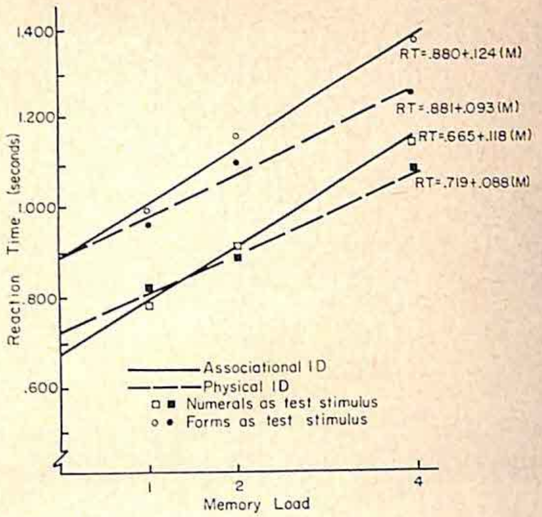


FIG. 1. RT memory load functions for associational and physical identity matches.

As indicated above, the slopes obtained in the present experiment when the format of the test stimulus was uncertain on each trial were close to the slopes obtained for both two-digit numerals and random forms when the format of the test stimulus was certain (and the same) on each trial. But the intercepts of the RT functions differed substantially: 360 msec. versus 719 msec. for numerals and 602 msec. versus 881 msec. for random forms. This comparison suggests that the introduction of format uncertainty drastically affected (lengthened) Stage 1 (encoding and preprocessing) time and had essentially no effect on Stage 2 (memory comparison) time.

In the analysis of the negative response data, the main effects (*a*) type of match, $F\ (1, 24) = 23.9$; (*b*) format of test stimulus, $F\ (1, 24) = 85.1$; and (*c*) memory load, $F\ (2, 48) = 223.3$; were significant just as in the positive response data. The Type of Match × Memory Load interaction however, failed to obtain statistical significance, $F\ (2, 48) = 1.3$, indicating that the profiles of the RT memory load functions did not differ significantly in slope. This suggests that either (*a*) recoding did not take place in the associational identity condition or (*b*) that recoding took place in both the physical and associational identity conditions.

The memory functions for positive and negative responses separately in the physical and associational identity conditions are given below:

$\mathrm{RT\ (sec.)} = .800 + .090\ (M)$
Positive physical identity condition

---

[5] This study was conducted by Paul Noel at the University of Texas, Austin, to partially fulfill requirements for a master's degree.

RT (sec.) = .830 + .108 ($M$)
Negative physical identity condition
RT (sec.) = .772 + .121 ($M$)
Positive associational identity condition
RT (sec.) = .865 + .119 ($M$)
Negative associational identity condition.

If the difference in slope for positive and negative responses is taken for these four functions, a difference of 2 msec/item is obtained for the associational identity condition, while a difference of 18 msec/item is obtained for the physical identity condition. Why should the positive and negative functions differ so much more for the physical identity condition than for the associational identity condition? Posner and Keele (1967) made an interesting assumption about negative responses in their same–different task: If a match cannot be made on the basis of physical identity, then a check is made on the basis of associational identity before a negative response is emitted. The data in the present experiment support this assumption. Apparently, if a match does not occur in the physical identity condition, a check involving recoding is made. Since recoding is a Stage 2 operation, this produces a change (increase) in the slope of the RT memory load function based on the negative responses for the physical as well as the associational identity conditions.

In the analysis of the error data, the primary trend was for error rate to increase with memory load, $F$ (2, 48) = 27.6, $p < .01$. The Type of Match × Memory Load interaction was significant, $F$ (2, 48) = 6.16, $p < .01$, as it was in the analysis of the RT data. Since the difference in RT across memory load for the associational and the physical identity conditions is the critical difference in the present experiment, the significant Type of Match × Memory Load effect in the error analysis is of concern because this possibly could modify the interpretation of the Type of Match × Memory Load effect in the RT data. However, as Table 1 shows, the difference in error rates for the associational and physical identity conditions is in the slope of the error rate-memory load function, and the increase in error rate across memory load is greater in the associational identity condition than in the physical identity condition. Thus, any correction of the RT data on the basis of errors would increase the slope of the RT memory load function more for the associational identity condition than for the physical identity condition. The differences in reaction time in the associational and physical identity condition, then, are not artifacts due to differential error rates; rather, the differential error rates have worked to make the difference in those two conditions seem smaller than they actually would be under ideal (equal error) circumstances.

## DISCUSSION

The difference in the physical and associational identity conditions was interpreted in the introduction as a recoding effect. The results of the present experiment localize this recoding effect in the slope of the RT memory load function. The slope of this function is interpreted as central processing time; thus when recoding was required for a match to occur, central processing time slowed from 90 msec/item (the slope of the RT memory load function for physical identity matches) to 121 msec/item (the slope of the RT memory load function for associational identity matches). It follows from the predictions made earlier on the basis of Sternberg's (1966) methodology, then, that the recoding effect is localized in Stage 2 of the Sternberg (1967) model of human information processing. This suggests that when the format of the encoded test stimulus representation differs from the format of the memorial representations of the one, two, or four positive set items, the *several* memorial representations are recoded into the format of the *single* test stimulus before the comparison at Stage 2 is made.

Why should the several memorial representations be recoded rather than the single test stimulus? As Sternberg (1967) has pointed out, identification of the test stimulus is not necessary for ascertaining the presence or absence of that stimulus in a memorized list of stimuli (the positive set). If the test stimulus were to be recoded, it would have to be identified (What stimulus has been presented?) before it could be recoded (What is the alternative representation of the identified stimulus?). Apparently, recoding the one, two, or four memorial representations is more efficient than identifying and recoding the single test stimulus. These data also indicate that a Node 2 match may not be a "name" match as suggested in Posner and Mitchell (1967) since naming would require identification of the test stimulus. For this reason, the label "associational identity" is preferred to the term "name identity."

This result could be due to the size of the stimulus ensemble used in the present experiment. It is well established (see Smith, 1968) that identification time is a function of stimulus uncertainty. Stimulus uncertainty in the present experiment was determined by the 16 two-digit numerals and the 16 random forms used as stimuli, and the size of the stimulus ensemble (stimulus uncertainty) remained constant across trials. Thus, the time it would take to identify the test stimulus should not vary across trials. Now, the uncertainty associated with the positive set did vary from trial to trial, and so recoding time should vary with $M$. Therefore, a comparison of the time it would take to identify the test stimulus and the time taken to recode the several memorial representations is not a comparison of the rate of recoding and the rate of identification. This comparison cannot be made in the present experiment because the rate of identification cannot be estimated.

The rate of recoding can be estimated, however, by comparing the central processing times for the recoding (associational identity) and no-recoding (physical identity) conditions. The difference in central processing time in the present experiment for those two conditions (see Fig. 1) was 31 msec/item of memory. The reciprocal of this difference (32 items/sec) is interpreted as the rate of recoding, and it is the same when random forms are being recoded as when numerals are being recoded.

If the memorial representations are recoded into the format of the test stimulus, the format of the test stimulus must be determined *prior* to recoding. For this reason, it is necessary to assume that the test stimulus is analyzed in Stage 1 to the extent that some global aspect—such as its format—is recognized. Neisser (1967) identified this type of gross analysis as preprocessing, and Sternberg (1967) has identified preprocessing as a Stage 1 operation. Also, since recoding takes place only on the trials when the formats of the memorized positive set and the test stimulus differ, one can assume that a format comparison takes place to ascertain whether the formats are the same or whether they differ. This would be a Stage 1 (preprocessing) operation.

There are alternative explanations of these data, of course.[6] One is based on the assumption that the test stimulus is recoded but only *after* a scan of the memory items has been attempted and has failed using the original (physical) representation of the test stimulus. Then another scan of memory would be attempted using the recoded version of the test stimulus (thus making it a Stage 2 effect), and this second scan would be required on associational identity trials and on negative response trials. This would account for the observed differences in slope. A similar explanation is based on the hypothesis that *dual* representations of the test stimulus are formed upon presentation, and that the representation in the same format as the memory items is tried first during the comparison process. If that fails, then the alternative representation is tried. Again, a second scan of the memory items would be necessary on associational identity and negative response trials, and this would account for the increase in slope observed for the associational identity and negative response conditions. Both of these explanations are viable, but both are based on the assumption that the test stimulus is identified before scanning takes place. This makes them less appealing than the earlier explanation offered because of the arguments presented above concerning identification of the test stimulus prior to the comparison process and because it is inconsistent with previous data in the memory scanning literature (see Sternberg, 1967).

The above results suggest that Stage 1 accomplishes not only stimulus encoding, but also a preprocessing function of format recognition and comparison. Thus, there are two outputs from Stage 1 which become inputs to Stage 2: the encoded version of the test stimulus and the format information. Before either recoding or comparison can take place, the format information must be processed to determine whether the memory items are in the appropriate format for comparison to the encoded, but still unknown, stimulus. This interpretation follows Posner and Mitchell's (1967) suggestion that the TOTE (test-operate-test-exit) concept of Miller, Galanter, and Pribram (1960) be used to interpret RT data in human information-processing tasks. The TOTE unit involves a comparison between a specified goal and the present state of the organism. The goal in this case is to have the test stimulus representation and the memory representations in the same format. If a test is made (a format comparison) and it fails, then recoding takes place. Since recoding is an operation which results in the goal being attained, this routine (recoding) is exited and a new routine (central comparison) is entered. Central comparison, then, would consist of comparing the test stimulus representation to the memorial representations (which are always in the same format due to the Stage 1 operations) to ascertain whether the test stimulus is one of the memorized items composing the positive set.

---

[6] The first author would like to thank Richard Wilton for suggesting several alternative explanations of these data including the two presented herein.

It should be emphasized that the data from the present experiment suggest that *S*s did *not* recode random forms (presented as positive set items) *upon presentation* into the provided name (two-digit numeral) code and then use those recoded representations of the positive set stimuli during the matching process when the test stimulus was presented. This assumption has been made by Klatsky and Atkinson (1970) and Klatsky, Juola, and Atkinson (1971) concerning data obtained in their experiments which had conditions similar to the NN and NF conditions of the present study. The addition of the FF and FN (see Table 1) conditions provides additional relevant data. If the strategy suggested above had been adopted, the presentation format would have been irrelevant since functionally the memorized positive set items would have always been in a name code. Then, recoding on test trials would have been necessary only (and always) when the test stimulus was a random form and never when the test stimulus was a two-digit numeral. This clearly was not the case since (*a*) the recoding effect was obtained in the associational identity condition when a name was presented as a test stimulus and (*b*) the recoding effect was not obtained in the physical identity condition when a form was presented as a test stimulus.

These results are consistent with the data from a cross-modality study reported by Chase and Calfee (1969, Exp. I). In that study, letters were presented to *S*s either auditorily or visually. Then an auditory or visual test stimulus was presented. The associational identity condition occurred when the positive set items and the test stimulus were presented in different modalities, and the physical identity condition obtained when they were presented in the same modality. The results of the Chase and Calfee study showed that the difference in the cross-modality (association identity) condition and the homogeneous-modality (physical identity) condition was in the slope of the RT memory load function. This indicates that a difference in format (as in the present study) and a difference in modality (as in the Chase & Calfee study) have a similar effect on the functions relating RT to memory load.

Originally (see Posner & Keele, 1967; Posner & Mitchell, 1967), an integral part of Posner's "nodes of processing" theory was that the memorial representation of a visually presented stimulus was recoded to and preserved in a name code. This assumption was based on research by Sperling (1960) and by Averbach and Coriell (1961) which identified "iconic" memory (see Neisser, 1967) as extremely labile and short-lived. Visually presented stimuli were assumed to be recoded and stored in a name code because visual information presumably is subject to a fast rate of decay and is lost within a very short time (less than 1 sec.).

In several studies (Chase & Calfee, 1969; Posner et al., 1969; Posner & Mitchell, 1969) it has been assumed that the Node 2 or name code is auditory in nature and, as a result, cross-modality experiments were conducted to investigate such associational identity matches. Even Tversky (1969), who used exclusively visual presentation of stimuli, assumed that certain kinds of visually presented stimuli (verbal stimuli) are dealt with in an auditory code, while other kinds of visually presented stimuli (pictorial stimuli) are processed within a visual code. The data from several recent studies show the inadequacy of the viewpoint that associational identity matches are made on the basis of auditory representations of the stimuli being compared. For example, Posner et al. (1969) and Tversky (1969) have shown that visual information can be preserved (or even generated) by *S*s who expect a visually presented test stimulus.

In the present experiment, there were relatively long intervals between presentation of a potentially matching target item and the test item. A minimum interval of 5 sec. occurred in the $M = 1$ condition, and the maximum interval of 12.5 sec. occurred in the $M = 4$ condition when the first item presented as a positive set item was the matching item. Despite these long intervals, the data in the present study suggest that *S* used a visual code to store or maintain the positive set items in "active" (Stage 2) memory. The rationale for this conclusion is provided by Posner et al. (1969). As they put it: "If only the name of a letter is present in store, it should make no difference whether or not the new letter has the same physical form. On the other hand, if full visual information was used in the match, reaction time should be much faster when the input matches the stored code in physical form [p. 1]." Since a recoding effect was observed both when random forms composed the positive set and when numerals (names) composed the positive set, one can conclude that "full visual information" was used in the matching process in the physical identity condition for names as well as for random forms. The data from the generation experiments of Posner et al. (1969) and Tversky (1969) suggest the recoding that takes place in the associational identity condition consists of generating the alternative *visual* representation of the stimulus in memory. Thus, it is suggested that in the present experiment, visual information was used in the comparison process both in physical and associational identity conditions. It is not suggested, however, that this visual code is "iconic" in nature. Rather, it is suggested that the labile iconic representation of a visually presented stimulus can be recoded into a longer lasting visual code as well as into a long-lasting auditory code as originally suggested by Sperling (1967). Results from a similar experiment have led Cruse and Clifton (1971) to make the same suggestion concerning visual information processing.

It should be noted further, however, that during a postexperiment debriefing, *S*s reported using the "names" (two-digit numerals) and an auditory rehearsal strategy to maintain the positive set stimuli in active memory. Apparently, as Sternberg (1967) suggests, an auditory rehearsal strategy can be used by *S*s to maintain or preserve the availability of visual representations of stimuli in active (not iconic) memory. Despite this auditory rehearsal strategy, the data of the present experiment indicate that the modality (or code) associated with the comparison process was visual.

## REFERENCES

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, **40**, 309–328.

BRIGGS, G. E. *Reaction time in human information processing*. (Tech. Rep. 69-5) Columbus, Ohio: Ohio State University Computer and Information Science Research Center, 1969.

BRIGGS, G. E., & BLAHA, J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, 1969, **79**, 385–402.

CHASE, W. G., & CALFEE, R. C. Modaility and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

CRUSE, D., & CLIFTON, C. *Recoding strategies and the retrieval of information from memory*. (Tech. Rep. No. 71-1) Amherst, Mass.: Cognitive Processes Laboratory, University of Massachusetts, 1971.

KLATSKY, R. L., & ATKINSON, R. C. Memory based on alternative test stimulus representations. *Perception & Psychophysics*, 1970, **8**, 113–117.

KLATSKY, R. L., JUOLA, J. F., & ATKINSON, R. C. Test stimulus representation and experimental context effects in memory scanning. *Journal of Experimental Psychology*, 1971, **87**, 281–288.

MILLER, G., GALANTER, E., & PRIBRAM, K. *Plans and the structure of behavior*. New York: Holt, 1960.

NEISSER, U. *Cognitive psychology*, New York: Appleton-Century-Crofts, 1967.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79** (1, Pt. 2).

POSNER, M. I., & KEELE, S. W. Decay of visual information from a single letter. *Science*, 1967, **158**, 137–139.

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392–409.

SMITH, E. E. Effects of familiarity on stimulus recognition and categorization. *Journal of Experimental Psychology*, 1967, **74**, 323–332.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

SPERLING, G. The information available in brief visual displays. *Psychological Monographs*, 1960, **74** (11, Whole No. 498).

SPERLING, G. Successive approximations to a model of short-term memory. *Acta Psychologica*, Amsterdam, 1967, **27**, 285–292.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

STERNBERG, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, **30**, 276–315. (a)

STERNBERG, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457. (b)

SWANSON, J. M., & BRIGGS, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, **81**, 223–239.

TVERSKY, B. Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, 1969, **6**, 225–233.

VANDERPLAS, J. M., SANDERSON, W. A., & VANDERPLAS, J. N. Statistical and associational characteristics of 1100 random shapes. *Perceptual and Motor Skills*, 1965, **21**, 414.

(Received June 5, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 10–20

# UNIVERSALS IN COLOR NAMING AND MEMORY[1]

ELEANOR ROSCH HEIDER[2]

*Brown University*

Four studies investigated the hypothesis that there are salient areas of the color space (here called "focal colors") which are universally the most linguistically "codable" and the most easily remembered. Experiment I completed previous location of these focal colors on the dimensions of hue, value, and saturation. In Exp. II, speakers of 23 different languages named a sample of focal and nonfocal colors. In Exp. III, 20 English speakers and 21 New Guinea Dani, speakers of a language which lacks hue names, remembered and recognized focal and nonfocal colors. In Exp. IV, 19 Dani learned names for focal and nonfocal colors in a paired-associates task. The results were that focal colors were given the shortest names and named most rapidly across languages, were the most accurately recognized both by English and Dani speakers, and could be paired with names with fewest errors. On the basis of these findings, linguistically causal interpretations of earlier language-cognition studies using color were challenged.

The color space has been considered an ideal domain for language-cognition research because it has been believed to be a domain of continuous, uniform physical variation which languages could arbitrarily "cut up" into different color name categories (Brown & Lenneberg, 1954; Lenneberg, 1967; Stefflre, Castillo Vales, & Morley, 1966). In fact, however, the results of several previous studies could equally have been interpreted as showing that some areas of the color space have a particular "distinctiveness."

The basic language-cognition hypothesis is that differences in the way natural languages code a domain are reflected in the way in which speakers of those languages "conceive reality" in that domain (cf. Whorf, 1950). Psychological language-cognition studies have attempted to correlate linguistic differences with specific nonlinguistic behaviors that could serve as an independent measure of cognition; the major group of studies obtaining positive results have used color as the nonlinguistic domain, color naming as the linguistic variable, and recognition memory as the cognitive measure.

Brown and Lenneberg (1954) found that the "codability" of a color (a composite measure of agreement in naming, length of name, and response latency in naming) correlated with memory accuracy (the accuracy with which a color could be remembered in a recognition task). A reasonable interpretation (cf. Krauss, 1968) has been that the more codable colors were better remembered, particularly under conditions of delayed recognition, specifically because of the greater ease with which they could be linguistically coded and stored in memory. However, the correlation between codability and memory did not hold true for all colors. Brown and Lenneberg's stimulus colors were all highly saturated, colors from the "outer shell" of the color space. The correlation between codability and memory failed to be replicated when the task was performed using Farnsworth-Munsell colors, an array of uniform low saturation (Burnham & Clark, 1955; Farnsworth, 1949; Lenneberg, 1961). Lenneberg attempted to explain this nonreplicability by the argument that the particular array used by Brown and

---

[1] This research was supported in part by the cross-cultural study of ethnocentrism, in part by the Foundations Fund for Research in Psychiatry Grant G67-392, and in part by funds made available to the Child Study Center of Brown University from the Grant Foundation and from the National Institute of Neurological Diseases and Blindness, National Collaborative Project (PH-43-68-13). A portion of the data was reported at a meeting of the Eastern Psychological Association, New York, April 1971. The assistance of Paul Kay is gratefully acknowledged.

[2] Requests for reprints should be sent to Eleanor R. Heider, who is now in the Department of Psychology, University of California at Berkeley, Berkeley, California 94720.

Lenneberg (1954) happened to include a few easily named colors mixed with many innominate colors, thus allowing "namability" to function as a distinctive cue. However, it is a factual matter that it is impossible to balance the number of easily named and innominate colors in a highly saturated array of Munsell colors, such as Brown and Lenneberg's, without using namable colors that are perceptually closer to each other than the innominate colors. The distinctiveness of codable colors in a saturated array representative of the color space appear to be an inherent, not accidental, aspect of such an array.

In apparent contradiction, a second linguistic variable—communication accuracy—was found to correlate with memory regardless of stimulus array (Lantz & Stefflre, 1964; Stefflre et al., 1966). This line of research, although quite relevant to the relation between interpersonal and intrapersonal communication, is not, however, relevant to the relation between a linguistic domain and the nonlinguistic domain which it encodes (Lenneberg, 1967).

More explicit evidence for color distinctiveness was provided by the recent work of two anthropologists (Berlin & Kay, 1969). Using foreign students as informants, Berlin and Kay first elicited, with no color stimuli present, the basic color terms in the informant's native language, "basic" terms distinguished from secondary terms by linguistic critera. The informant was then asked to pick the best example of each of his basic color names from an array of all hues and values (brightnesses) in the *Munsell Book of Color*, each at maximum saturation available for that hue and value. (Note that this was not a coding task; informants were never asked to name color stimuli). Berlin and Kay found: (*a*) that there was a limited number of basic terms in any language, at most three achromatic and eight chromatic, and (*b*) that the chips designated as best examples of these basic terms tended to fall into "clusters." One cluster of best examples included chips close to the chip an English speaker gave as the best example of "red," another cluster close to the best example of "green," and similarly for each other "basic" color term ("yellow," "blue," "brown," "pink," "orange," and "purple"). Berlin and Kay called these best examples of color terms "focal points" and argued that the previous anthropological emphasis on cross-cultural differences in color names was derived from looking at boundaries of color names rather than at color-name focal points.

In language-cognition research, when correlations were found between an aspect of language and cognition, interpretations tended to credit the linguistic variable with a causative role in the cognitive performance. The finding of such correlations does not logically imply that there need be cross-cultural differences; however, if both variables were invariant across languages, the interpretation that the relation showed an effect of language on thought would become implausible. The hypothesis of the present research is that just such an invariance exists: that there are specific areas of the color space—defined as particular intersections of the three dimensions hue, value, and saturation—which are universally the most codable and the most accurately remembered (both in short- and long-term memory) and that it is these areas that form the focal points of basic color names across languages.

The first task was to locate the hypothesized distinctive areas of the color space; presumably these would be the areas most central in the cluster of colors given as the best example of basic color names. Berlin and Kay (1969) mapped the location of focal areas on the dimensions of hue and value but did not do so for saturation. Experiment I of the present research was designed to establish the location of focal colors on the saturation dimension.

The second question was whether the focal colors located by the first experiment were more codable (i.e., were given shorter names and named more rapidly) than nonfocal colors across languages. Berlin and Kay's (1969) data suggest this as a reasonable hypothesis, but, because they did not have their informants code colors, do not demonstrate it. That is, Berlin and Kay's data would only constitute proof of the greater codability of focal colors if pointing to the referent of a name and producing a name for a referent were the same process; clearly they are not. Furthermore, a pilot study showed that American college students tended to produce rather elaborate encodings for all colors, including those which they themselves later designated as the best examples of basic English color names; e.g., focal red received names such as "scarlet," "medium-dark red," and "true red." Experiment II was designed to test empirically the codability of focal and nonfocal colors across languages.

The third question was whether focal colors would be more accurately remembered than nonfocal colors. This question is only of limited interest if tested with speakers of English or other languages possessing the full complement of basic color terms; the interesting problem is whether focal colors are more memorable even to speakers of a language in which these colors do not represent the best examples of basic color terms. The language of the Dani of New Guinea lacks basic chromatic color terms (K. G. Heider, 1970) and, therefore, through a cross-cultural comparison, allowed a test of the question. Experiment III used a short-term recognition task similar to that employed by Brown and Lenneberg (1954) to test memory for focal and nonfocal colors with English speakers and with a sample of New Guinea Dani. The hypothesis of Exp. IV was that focal colors would be most easily remembered in association with color names in a long-term memory task—of course, by the speakers of the language which initially lacked color names.

A secondary concern, in all four experiments, was whether the focal colors differed among themselves in codability or memorability. From their search of the anthropological literature, Berlin and Kay (1969) argued that color terms in all languages evolved in the following fixed order: (*a*) black-white, (*b*) red, (*c*) green-yellow, (*d*) blue, (*e*) brown, (*f*) pink-orange-purple-gray. A strong hypothesis of differences among focal colors is that linguistic and cognitive measures will conform to that order. A less demanding hypothesis, and one that is made quite reasonable from the finding of physiologically "primary" hues (De Valois & Jacobs, 1968) is that the four primary chromatic focal areas (those for red, yellow, green, and blue) will prove more codable and memorable than the other chromatic focal areas.

## EXPERIMENT I

The purpose of the first experiment was to determine, for the dimension of saturation, the location of the focal points of basic color names.

### *Method*

*Subjects.*—Two groups of *S*s were employed: a group of native English speakers and a mixed language group composed of available informants from other languages. The English-speaking *S*s were 20 college undergraduate girls; the non-English group was composed of 10 foreign students. The languages represented were: Hungarian, Housa, Italian, Korean, Japanese. Javanese, Navaho, Portugese, Spanish, and Tagalog. All foreign *S*s were somewhat bilingual in English. The *S*s were screened for color blindness. All *S*s were paid volunteers.

*Stimuli.*—In this and all subsequent experiments, Munsell colors of glossy finish were used and are referred to in Munsell notation. The International Commission on Illumination tristimulus values and chromaticity coordinates are obtainable from Nickerson, Tomaszewski, and Boyd (1953).

In all Unites States experiments, stimuli were illuminated from above by a General Electric standard daylight fluorescent lamp; Dani were tested in daylight.

The first step in choosing stimulus colors was the location of the most "focal" color chip in the array used by Berlin and Kay (1969) for each of the eight basic color-name focal areas. This was done by finding the geometric centers of the irregular areas formed by each cluster of best-example choices, then by picking the chip, touching that central point, which had received the highest number of best-example choices for that cluster. The chips chosen were: red 5R 4/14; yellow 2.5Y 8/16; green 7.5G 5/10; blue 2.5PB 5/12; pink 5R 8/6; orange 2.5YR 6/16; brown 5YR 3/6; purple 5P 3/10. These chips were all at maximum saturation for that hue and value provided in the *Munsell Book of Color*. To these chips were added all chips of lesser saturation of the same hue and value. The number of chips added varied according to the saturation of the original; the range was from 3 each, for pink and brown, to 8 each, for yellow and orange. The final array was composed of 53 chips.

*Procedure.*—Chips were mounted in separate 5.08 × 5.08 cm. white cards arranged in a randomly ordered array on a gray surface (Munsell value/5).

Each *S* saw a different random order. Instructions were the following: "Out of all these colors, find the one that, to you, is the best example of 'X'." "X" was, in turn, each of the basic color names in *S*'s language. When foreign students were tested, the basic color names of *S*'s language were elicited before testing began. The *S*s were seen individually.

### *Results*

For the United States *S*s, 95% of all choices of best examples of basic color names fell on chips of the two highest saturation levels. That is, of the 160 choices of a best example of a basic color name made by United States *S*s (20 *S*s making 1 choice for each of eight basic color names), only 8 of those choices failed to be a chip of the two highest saturation levels available for the hue–value to which the name referred. For the foreign *S*s, 92.5% of the choices were of the chips of the two highest saturation levels. That is, of the total of 80 choices made by the 10 foreign *S*s, only 6 of those choices failed to be of the two most saturated chips available for the hue–value to which that name referred. It was quite clear, without further analysis, that the most saturated colors were the best examples of basic color names both for English speakers and for speakers of the other 10 languages represented.

## EXPERIMENT II

The basic hypothesis of the second experiment was that the same colors, specifically the focal points of basic color terms located by Berlin and Kay (1969) and by the results of the first experiment, were the most codable colors across languages. Only two measures of codability were used here, length of name and response latency of naming, because sufficient numbers of speakers of each non-English language were not available for the use of inter-*S* agreement as a measure. A secondary hypothesis was that focal colors would differ among themselves in codability; the strong form of this hypothesis was that codability scores would follow Berlin and Kay's (1969) evolutionary order of color names, the weak form that the four primary basic chromatic colors would be more codable than the other four focal colors.

### *Method*

*Subjects.*—The *S*s were 23 adult native speakers of a language other than English—13 males, 10 females—each a speaker of a different language. Fourteen of the *S*s were tested in the United States, test instructions presented in English; 9 *S*s were tested in Indonesia, test instructions presented in Bahasa Indonesia. The languages, grouped into language families, were: *Indo-European*—German, Italian, Polish, Swedish, Hindi, Urdu; *Austronesian*—Tagalog, Malay, Bahasa Indonesia, Javanese, Batak, Sundanese; *Sino-Tibetan*—Thai, Burmese, Cantonese, Mandarin; *Afro-Asiatic*—Yoruba, Housa, Arabic, modern Hebrew; *Miscellaneous*—Hungarian, Korean, Japanese. The *S*s were screened for color blindness. No *S* had served in Exp. I. All *S*s were paid volunteers.

*Stimuli.*—Three classifications of colors were used as stimuli: (*a*) Focal colors consisted of the eight chromatic chips representing the center of each best-example cluster, with the addition of black (N1/) and white (N9/). The chromatic chips were listed in Exp. I. (*b*) Internominal colors were chips chosen from the centers of those areas in which no chips had been designated as the best example of a basic color name by any language. The attempt was made to choose internominal colors matching the focal colors as closely as possible in either hue, value, or saturation. There were eight internominal areas on the Berlin and Kay (1969) array; the colors chosen from those areas were: 2.5YR 8/6; 10Y 6/10; 10Y 4/6; 10GY 8/10; 10BG 7/6; 10BG 4/6; 5P 7/6; 5RP 3/10. (*c*) Boundary colors were colors adjacent to a focal cluster on one side with an internominal area or array border on the other side. Criteria for choosing boundary colors were that the chip be two or more spaces distant from the focal chip of that cluster, that the chip be directly adjacent to a chip that had been chosen as a best example by at least one informant, and that the chip had not itself been chosen as a best example by any informant. Thirteen chips met these criteria and were included: 2.5R 9/2; 5R 2/8; 2.5YR 7/10; 7.5YR 8/8; 2.5Y 3/4; 7.5Y 9/2; 10GY 7/8; 10G 3/4; 7.5B 6/8; 10B 3/6; 7.5PB 2/10; 5P 6/8; 10P 7/6.

*Procedure.*—Each chip was mounted individually on a 7.62 × 12.70 cm. white card. Cards were shown one at a time to each *S* who was instructed to write, in his own language, what he would call each color. The cards were "shuffled" after each testing so that each *S* saw the colors in a different order. Response latency, measured by a stopwatch, was the interval of time between *E*'s initial presentation of the color and *S*'s initial movement of the pencil in writing the answer. To control for *E* bias, most testing was done by an *E* who did not know the classification of colors or the hypotheses upon which the experiment was based, and *E*'s face was obscured from *S* during testing. If *S*'s language used a non-Roman orthography, *S* was asked, at the conclusion of testing, to transliterate the answers into Roman orthography.

TABLE 1

CODABILITY OF FOCAL, INTERNOMINAL, AND BOUNDARY COLORS

| Dependent variable | Stimulus colors | | |
|---|---|---|---|
| | Focal | Internominal | Boundary |
| No. of words | 1.82 | 2.59 | 2.77 |
| No. of letters[a] | 8.14 | 12.20 | 12.68 |
| Response latency[b] | 9.9 | 14.8 | 13.1 |

[a] Analyses performed using $\log_{10}$ of number of letters.

[b] Latency means shown in seconds. All analyses were performed using $\log_{10}$ of time in seconds.

*Results*

Table 1 shows the mean length of name and response latency of naming for focal, internominal, and boundary colors. Length of name was determined by two separate measures: the mean number of words used to name the color and the mean number of letters. For languages not written in Roman orthography, the length measures were derived from the transliteration. Significance of the length of name results was analyzed in two ways. The raw scores for both name variables were skewed toward the lower end of the range. A $\log_{10}$ transformation of the number of letters per encoding rendered the distribution of that variable reasonably close to a normal curve (a $\chi^2$ of the departure of the percentile frequency of the distribution of the transformed scores from the distribution expected from a normal curve was not significant, $\chi^2$ (9) = 4.82). Number of letters was, therefore, analyzed with a one-way analysis of variance, *S*s × Treatments; each category of color (focal, internominal, and boundary) constituted a treatment. The difference in length of name between the three categories of color was significant, $F$ (2, 22) = 8.41, $p < .01$. Tests for correlated measures between the categories of colors showed that focal colors were given significantly shorter names than internominal colors, $t$ (22) = 3.39, $p < .01$, that focal colors were given significantly shorter names than boundary colors, $t$ (22) = 4.05, $p < .01$, but that internominal and boundary colors did not differ significantly from each other, $t$ (22) = .58.

The number of words per name remained skewed even with the $\log_{10}$ transform and was, therefore, analyzed with the nonparametric sign test (Siegel, 1956). Of the 23 *S*s, 19 produced a smaller mean number words for focal than for internominal colors (with one tie) and 21 produced a smaller mean number of words for focal than for boundary colors. These proportions would be expected by chance in a binominal distribution, $p < .001$. Fourteen of the *S*s produced a greater mean number of words for boundary than for internominal colors; for 9 *S*s the opposite was the case, a difference which is not significant by the sign test. In summary, for both measures of naming, focal colors were given shorter names than either internominal or boundary colors; however, internominal and boundary colors did not differ from each other.

Latency of naming was analyzed by the same procedures as letters per name. Results were similar to the results for the length of name measures. The difference in naming latency between the three categories of color was significant, $F$ (2, 22) = 4.52, $p < .05$. Tests for correlated measures between all pairs of categories of color showed that focal colors were named more rapidly than internominal colors, $t$ (22) = 4.74; that focal colors were named more rapidly than boundary colors, $t$ (22) = 4.67, both $p$s $< .01$; but that internominal and boundary colors did not differ from each other, $t$ (22) = 1.50.

The results for both measures of codability (length of name and naming latency) were consistent. Focal colors were given shorter names and named more rapidly than nonfocal colors. Internominal and boundary colors did not differ significantly from each other. As had been found by Brown and Lenneberg (1954), the two measures of codability were correlated with each other across colors: $r = .57$, $p < .01$ for $r < 0$.

The languages included in the study do not represent all of the world's language families but do represent considerable diversity of linguistic stock. Since four language families were represented by more than one language (Indo-European 6, Austronesian 6, Sino-Tibetan 4, Afro-Asiatic 4), it was possible to test whether the results differed significantly between language families. For this purpose a two-way analysis of variance, mixed model, was performed: category of color constituted the within-*S* variable, language family the between-*S* variable. Neither the between-*S* main effect due to language family nor the interaction approached significance, main effect, $F$ (1, 21) = .77; interaction, $F$ (2, 42) = 1.93.

Inspection of the data for individual *S*s revealed additional facts. Individual length and latency means for focal and nonfocal colors tended to be in the same direction as the group means; in fact, for 19 out of the 23 *S*s, the focal and nonfocal color means for all three variables were in the same direction as the group means. Furthermore, as had been suggested by the pilot study on English speakers, the length difference between focal and nonfocal colors was not simply the result of *S*s naming focal colors with basic color names and nonfocal colors with secondary names. The *S*s showed consistent individual differences in naming elaboration; some used one- or two-word color names for most chips, while others used long descriptive phrases for most chips (the correlation across *S*s between number of letters in the focal and nonfocal names was significant: $r = .51$, $p < .01$). However, within their style preference, most *S*s consistently used shorter words or phrases for the focal than for the nonfocal colors.

A second hypothesis concerned codability differences within the focal colors. The codability order of colors, from most to least codable, was as follows: (*a*) *number of words*—black, white, yellow, purple, brown, red, pink, blue, orange, green; (*b*) *response latency*—black, yellow, white, purple, blue, red, pink, brown, green, orange. Except for the fact that the two achromatic colors "black" and "white" occurred early in the lists, neither order "looked like" the order in which Berlin and Kay (1969) claim color terms evolve, nor were the rank-order correlations of either order with the "evolutionary" order significant.

The weaker form of the order hypothesis did receive partial confirmation. While neither measure of length of name showed a significant difference (although both were in the predicted direction), primary focal colors were named significantly more rapidly than the nonprimary focal colors; $t$ (22) = 2.86, $p < .01$ (for correlated measures).

## Experiment III

Language-cognition studies with color have typically employed a short-term recognition task as the memory measure (Brown & Lenneberg, 1954; Lantz & Stefflre, 1964; Stefflre et al., 1966). The basic hypothesis of the third experiment was that in such a task, focal colors can be remembered more accurately than nonfocal even by speakers of a language that lacks basic hue terms and in which, therefore, focal colors do not represent the best examples of basic color names. Such a language is spoken by the Dani of Indonesian New Guinea, a stone-age people previously described by K. G. Heider (1970) as having only two color terms. These divide the color space on the basis of brightness rather than hue. Color systems of that character have been reported for other cultures as well and form Stage I, the first and simplest stage, of Berlin and Kay's proposed evolutionary ordering of color systems. Data from a study verifying the nature of Dani color terminology and tables showing the distribution of the two basic Dani color terms over arrays of Munsell chips are available in Heider and Olivier (1972).

Experiment III was designed in two parts: (*a*) An English-speaking group was tested; this was intended primarily as a replication of Brown and Lenneberg's (1954) findings using the present systematically derived focal and nonfocal colors. (*b*) A Dani group was tested; this was intended to test the hypothesis that memory differences between focal and nonfocal colors exist even when codability differences do not. The two groups were given identical tasks so that possible interactions between culture and stimulus type

TABLE 2

Accuracy and Latency of Color Memory

| Dependent variable | Stimulus colors | | |
|---|---|---|---|
| | Focal | Internominal | Boundary |
| American *S*s | | | |
| No. correct | 5.25 | 3.22 | 2.51 |
| Correct response latency[a] | 6.0 | 9.8 | 12.5 |
| Incorrect response latency[a] | 12.4 | 13.5 | 11.8 |
| Dani *S*s | | | |
| No. correct | 2.05 | .47 | .71 |
| Correct response latency[a] | 2.5 | 2.9 | 3.6 |
| Incorrect response latency[a] | 3.6 | 3.5 | 3.6 |

[a] Latency means shown in seconds. All analyses were performed using $\log_{10}$ of time in seconds.

could be compared. Secondary hypotheses in Exp. III were equivalent to the hypotheses concerning within-focal-color ordering of Exp. II.

### *Method*

*Subjects.*—The *S*s were 20 native English speakers —12 females, 8 males—and 21 Dani, all females. All *S*s were paid volunteers and were screened for color blindness. Dani *S*s were monolingual and were screened for the use of color names; only Dani whose color-name usage was restricted to the two basic terms "mili" (roughly "dark") and "mola" (roughly "light") were used. In naming colors, such Dani *S*s tended to "chant" the two names at a constant rate so that neither length of name nor gross latency differences existed between colors for these *S*s.

*Stimuli.*—Two sets of colors were used in the memory task, colors *S* was to remember and an array of colors from which *S* was later to distinguish the memory colors. A pilot study with United States *S*s tested the chromatic focal, internominal, and boundary colors listed in Exp. II as the memory colors and the 320-color chromatic array used by Berlin and Kay (1969) as the array from which they were recognized. The task, so constructed, proved too difficult even for American college students; there were few completely accurate memory choices, and *S*s found the test extremely frustrating. A smaller array from which the memory colors were to be distinguished was constructed by eliminating Levels 2.5 and 7.5 of each hue from the Berlin and Kay array, a change which also made the hue and value steps more nearly perceptually equal. The resultant 160-color array was mounted on white cardboard in "Munsell order," hues horizontally, values vertically. Changes in the array necessitated some changes in the memory colors; chips of 2.5 and 7.5 hues were replaced with an immediately adjacent 5- or 10-hue chip which met the criteria for that color category described in Exp. I and II. The final memory stimuli consisted of all of the 5- and 10-hue chips in each category used in Exp. II, with the addition of the following replacements for the 2.5 and 7.5 chips: focal color replacement—10YR 8/14, 5YR 7/Max, 10GY 5/12, 10B 4/10; internominal color replacement—5YR 8/6; boundary color replacements—5Y 3/4, 10Y 9/6, 5B 6/8. An additional effect of this method was to equalize the number of colors (to eight) in each category.

*Procedure.*—Procedures were essentially identical for United States and Dani *S*s. Test chips were mounted on 7.62 × 12.70 cm. white cards. The *S* was shown a single test chip for 5 sec., waited 30 sec., and was then shown the entire 160-chip array and asked to select from it the chip he had seen. The 30-sec. delay was not filled by any activity. The array was covered with white cardboard during the presentation of the chip and the wait. Test chips were presented to *S* in random order; after each testing, the cards were shuffled so that each *S* saw the colors in a different order.

Response latency, measured by stopwatch, was the interval between *E*'s uncovering of the array and *S*'s pointing to one of the array chips as his answer. As in the second experiment, the attempt was made to control for *E* bias. In the United States, testing was performed by an *E* who did not know the classification of colors nor the purpose of the experiment. Although Dani *S*s were tested by the author, the testing was so arranged that only *E*'s hands were visible to the *S* during testing.

### *Results*

The basic memory accuracy score was *S*'s number of correct recognitions of focal, internominal, and boundary colors. Table 2 shows the means of these scores. Results were analyzed by a two-way analysis of variance, mixed model (a Type 1 mixed model, Lindquist, 1953). Categories of colors were the within-*S* variable, culture the between-*S* variable. Both main effects were significant: United States *S*s were more accurate than Dani, $F\ (1, 40) = 15.07$; and the difference between categories of color was significant, $F\ (2, 78) = 6.11$, both $p\text{s} < .01$. The interaction was not significant, $F\ (2, 78) = .91$. Tests of the simple effects of the within-*S* variable at each level of the between-*S* variable (the effect of categories of color within each culture) were also performed. Because it could not be assumed that the error (within cells) was homogeneous for both cultures, the analysis corresponded to a simple Treatments × *S*s design performed within each culture (Lindquist, 1953). The difference between categories of color was significant within the United States culture, $F$ (2, 19) = 6.39, and within the Dani culture, $F$ (2, 20) = 6.84, both $p$s $< .01$. Tests for correlated scores between all pairs of categories of color showed that within the United States culture, focal colors were remembered more accurately than internominal colors, $t$ (19) = 3.29; focal colors were remembered more accurately than boundary colors, $t$ (19) = 4.86, both $p$s $< .01$, but that internominal and boundary colors did not differ significantly from each other, $t$ (19) = 1.04. Within the Dani culture, the same pattern of results were obtained. Focal colors were remembered more accurately than internominal colors, $t$ (20) = 4.09, $p < .01$; focal colors were remembered more accurately than boundary colors, $t$ (20) = 3.74, $p < .01$; but internominal and boundary colors did not differ from each other, $t$ (20) = .97.

In summary, the results for memory accuracy were highly consistent; focal colors were remembered more accurately than nonfocal both by Americans and Dani. The difference between internominal and boundary colors was not significant for either culture. Not surprisingly, United States recognition accuracy was higher than Dani; however, it is important to note that there was no interaction between culture and treatment; that is, the direction of differences was the same within both cultures.

Several factors which might have biased these results were also analyzed. In the first place, Munsell chips were not perceptually equidistant; therefore, some chips in the memory array were more "discriminable" (perceptually more distant) from their neighbors than other chips. Brown and Lenneberg (1954) found that discriminability correlated with accuracy of recognition (though not with codability). It was possible, therefore, that discriminability was a confounding factor in the present study. Discriminability scores for the 24 test colors were computed by the method described in Brown and Lenneberg (1954). This is a method for computing the discriminability of a test chip from its neighboring chips in an array. In the present study, neighboring chips were defined as all those chips both one and two steps from the test chips in hue and brightness (in the memory array described previously). The score for each test color, thus, consisted of the combined scores of its discriminability from the 20 chips surrounding it in that particular array. The 4 immediately adjacent chips were weighted proportionally higher by a factor of 1. For these purposes, the array was considered circular with red adjacent to purple; otherwise, "edges" were treated as in Brown and Lenneberg (1954). Tests ($t$) between the discriminability scores thus derived were computed between focal, internominal, and boundary colors; none approached significance.

A second possible bias might have been introduced by a differential tendency of *S*s to guess focal or nonfocal colors when the correct color was not known. To test whether a correction for guessing was needed, the number of incorrect designations of focal and of nonfocal colors was compared by $t$ test for both cultures. In fact, guessing one test chip when another was correct was very rare and did not differ for focal and nonfocal colors in either culture. A final problem was presented by the small number of correct recognitions by Dani and the resulting restricted variance. The results were, therefore, checked by an analysis which used a different accuracy measure for which Dani variance was greater. Error distance scores were computed, which consisted of the mean distance between each chip and and all of the chips which had been incorrectly designated for that chip in incorrect recognition trials (distances between chips were computed from the revised spacing of Munsell chips in Newhall, Nickerson, & Judd, 1943). The same analyses were applied to this measure as to recognition accuracy; identical results were obtained from the two measures.

Response latencies were computed separately for correct and incorrect answers. The mean scores are shown in Table 2. The same mixed-model, two-way analysis of variance as described for the memory accuracy variable was applied separately to the two types of latency measures. For both measures, the between-$S$ main effect was significant; that is, Dani responded more rapidly than United States $S$s both when giving correct responses, $F$ (1, 40) $=8.78$, and when giving incorrect responses, $F$ (1, 40) $= 7.51$, both $p$s $< .01$. For latency of incorrect responses, none of the other effects nor any of the $t$ tests were significant. For latency of correct responses, the difference between the three categories of color was significant, $F$ (2, 78) $= 3.52$, $p < .05$, and the simple effect for differences between the three categories of color within the United States culture was significant, $F$ (2, 19) $= 4.22$, $p < .05$; however, the simple effect was not significant within the Dani culture, $F$ (2, 20) $= 1.77$. Within the United States culture, $t$ tests for correlated measures showed that correct responses were given more rapidly for focal than for internominal colors, $t$ (19) $= 2.47$, $p < .05$, and more rapidly for focal than for boundary colors, $t$ (19) $= 3.06$, $p < .01$, but that the difference between internominal and boundary colors was not significant, $t$ (19) $= 1.69$. For correct response latency, the interaction between culture and treatment was not significant, $F$ (2, 78) $= 1.61$.

A secondary question of the study concerned memory differences within the focal colors. The order of memory accuracy, from most to least remembered was: (*a*) *United States*—orange, yellow, red, blue-brown, green, pink-purple, and (*b*) *Dani*—orange, yellow, red, purple, green-pink, brown, blue. Neither order was significantly correlated with Berlin and Kay's proposed evolutionary order.

Accuracy and latency differences for the four primary and four nonprimary focal colors were analyzed. There were no significant differences in accuracy. However, as had been the case in Exp. II, there was a significant difference in latency. The latency of correct answers was shorter for the primary than for the nonprimary focal colors for the United States alone, $t$ (19) $= 3.94$, and for the United States and Dani combined, $t$ (40) $= 3.05$, both $p$s $< .01$.

## EXPERIMENT IV

The previous experiment was designed to be relevant to the findings of earlier language-cognition studies and so used short-term recognition memory as the cognitive variable. The hypothesis of Exp. IV was that focal colors were also more easily retained in long-term memory and can, thus, more easily become associated with color names.

### *Method*

*Subjects.*—The $S$s were 19 Dani—17 males, 2 females—screened for color blindness and color-name use as were Dani $S$s in Exp. III. No $S$ had served in Exp. III. The $S$s received payment only after they had completed the task to criterion (a mean of 3½ days testing). Five additional $S$s failed to return on successive days to continue training and were dropped from the study.

*Procedure.*—The $S$s were required to learn 16 stimulus-response pairs. The eight focal and the eight internominal colors described in the previous two experiments served as stimuli. Each color was paired with a separate response word. The response words used were all names of Dani *sibs* (a *sib* is a nonterritorial, unilineal descent group somewhat like a clan). Sib names were all well known to Dani and, as nearly as could be determined in pretesting, all approximately equally frequent, familiar, and meaningful. Furthermore, measures of association between sib names were easily obtainable (Dani readily listed sib names), and it was possible to use names with approximately equal interassociation value. In order to randomize effects of any uncontrolled memorability differences in sib names, each $S$ received a different set of pairings of colors and names. An $S$'s own sib name was never used as one of his responses.

The task was described to each $S$ as learning a new language that $E$ would teach him and for which he would be paid when learning was completed. He was warned that the learning could take several days. At the start of the first testing session, the colors (mounted on 7.62 × 12.70 cm. white cards) were laid in random order in front of $S$; and he was told the "name" of each, which he repeated after $E$. Thereafter, the cards were gathered into a pack, shuffled, and presented one at a time to $S$ who was required to say the name of the color. The $S$ received feedback after each response; he was praised if correct and told the correct response if incorrect. After each run the cards were reshuffled. The $S$s received five runs a day on successive days until the criterion of one perfect run had been achieved. A record was kept of the stimulus color and *S*'s response on each trial.

### *Results*

The mean number of errors per color for focal colors was 7.3, for internominal colors, 9.9. A $t$ test for correlated measures showed the difference significant, $t$ (18) = 3.27, $p < .01$. Focal colors did become associated with color names more rapidly than nonfocal colors.

The order in which the focal colors were learned, from lowest to highest errors, was: red, pink, green, orange, purple, blue, yellow, brown. This order did not correlate significantly with Berlin and Kay's proposed evolutionary order. The difference between errors for primary and nonprimary focal colors was significant, $t$ (18) $=1.69$, $p < .10$.

## GENERAL DISCUSSION

Clearly there exist linguistic codes that influence memory: one obvious example is the memory advantage gained from learning to chunk strings of zero and one digits into a higher order code such as binary or octal (cf. Miller, 1956). In a sense, the "language-cognition hypothesis" is that differences in codes analogous to "learning binary" exist between natural languages. To demonstrate that natural language codings of a domain are determiners of human memory for the nonlinguistic stimuli of the domain, evidence, such as the following, should be obtainable: (*a*) Codes for the stimuli should vary between languages. In the domain of colors, the particular colors to be most codable should be different in different linguistic communities, perhaps depending on the ecology of color in the particular culture. (*b*) Memory should, at the very least co-vary with changes in the code. The present research has shown that these conditions are not present in the domain of colors. The same colors were most codable in a diverse sample of languages (which violates *a*), and those colors were best remembered even in a culture where the linguistic factors of hue terms and codability were absent (which violates *b*).

The objection might be raised that *b* was not violated because the generally poor memory performance of the Dani did co-vary with lack of Dani hue names. However, there are many other differences between the Dani and American culture to which overall memory performance might better be attributed. Dani life contains neither school, work, nor interpersonal relations which produce overloads of information; thus, Dani have neither need for, practice with, nor any explicit training in the use of memory control processes. An interaction in memory between culture (Dani vs. U. S.) and treatment (focal vs. internominal vs. boundary colors) might have proved relevant to the present hypotheses; a main effect for culture alone (in the direction of Dani memory inferior to U. S.) cannot be so interpreted.

A second problem raised by the present research is the question of why color universals in naming and memory were not discovered by previous language-cognition studies. For example, Stefflre et al. (1966) found no correlation between the colors best remembered by speakers of Spanish and Yucatec. The answer is probably contained in the finding of Exp. I that focal colors are all of the highest saturations available for that hue and value. In order to "control" value and saturation, arrays of medium value and low saturation have typically been preferred. Stefflre et al. used the Farnsworth-Munsell colors, which are of low saturation and contain no focal colors. Holding value and saturation constant in an experimental array of colors does not eliminate effects of the particular level of value and saturation used, if those dimensions are significant.

Accepting, for the sake of argument, that focal colors are universally "salient," how might such saliency be related to universals in color naming? The superior retention of focal colors in long-term memory in the name-learning task demonstrated by Exp. IV suggests one mechanism by which the saliency of focal colors may lead to evolution of those colors as the core meaning of basic color terms across languages. Further evidence for the role of focal color areas in the development of color names in American children (Heider, 1971a) and in the experimentally controlled learning of color concepts (as opposed to specific associations) by Dani (Heider, 1971b) support the central role played by focal colors in learning the meaning of color names.

If the saliency of focal colors plays a role in the learning (and, by implication, the linguistic evolution) of basic color names, then it is not unreasonable to suppose that saliency differences among focal colors would lead to an evolutionary order in the development of color names. Although none of the measures of the present research supported the specific evolutionary order proposed by Berlin and Kay (1969), there was some confirmation (the significant response latency data of both Exp. II and III and the significant error data of Exp. IV) that Berlin and Kay's first four colors were more codable and memorable than the latter four. Such findings are compatible with the evidence that those four colors (red, yellow, green, and blue) may be the physiologically primary hues (cf. De Valois & Jacobs, 1968).

Given the attributes of focal colors—their occurrence as exemplars of basic color names, their linguistic codability across languages, and their superior retention in short- and long-term memory—it would seem most economical to suppose that these attributes are derived from the same underlying factors, most likely having to do with the physiology of primate color vision. In short, far from being a domain well suited to the study of the effects of language on thought, the color space would seem to be a prime example of the influence of underlying perceptual-cognitive factors on the formation and reference of linguistic categories.

## REFERENCES

Berlin, B., & Kay, P. *Basic color terms: Their universality and evolution.* Berkeley: University of California Press, 1969.

Brown, R. & Lenneberg, E. A study in lagnuage and cognition. *Journal of Abnormal and Social Psychology*, 1954, **49** 454–462.

Burnham, R. W., & Clark, J. R. A test of hue memory. *Journal of Applied Psychology*, 1955, **39**, 164–172.

De Valois, R. L., & Jacobs, G. H. Primate color vision. *Science*, 1968, **162**, 533–540.

Farnsworth, D. *The Farnsworth-Munsell 100 hue test for the examination of color discrimination: Manual.* Baltimore: Munsell Color Company, 1949.

Heider, E. R. "Focal" color areas and the development of color names. *Developmental Psychology*, 1971, **4**, 447–455. (a)

Heider, E. R. Natural categories. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **6**, 43–44. (Summary) (b)

Heider, E. R., & Olivier, D. The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, 1972, in press.

Heider, K. G. *The Dugum Dani: A Papuan culture in the highlands of West New Guinea.* Chicago: Aldine, 1970.

Krauss, R. M. Language as a symbolic process in communication: A psychological perspective. *American Scientist*, 1968, **56**, 265–278.

Lantz, D., & Stefflre, V. Language and cognition revisited. *Journal of Abnormal and Social Psychology*, 1964, **69**, 472–481.

Lenneberg, E. Color naming, color recognition, color discrimination: A re-appraisal. *Perceptual and Motor Skills*, 1961, **12**, 375–382.

Lenneberg, E. *Biological foundations of language.* New York: Wiley, 1967.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Newhall, S. M., Nickerson, D., & Judd, D. B. Final report of the OSA subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America*, 1943, **33**, 385–418.

Nickerson, D., Thomaszewski, J. J., & Boyd, T. F. Colorimetric specifications of Munsell repaints. *Journal of the Optical Society of America*, 1953, **43**, 163–171.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Stefflre, V., Castillo Vales, V., & Morley, L. Language and cognition in Yucatan: A cross-cultural replication. *Journal of Personality and Social Psychology*, 1966, **4**, 112–115.

Whorf, B. L. *Four articles on metalinguistics.* Washington: Foreign Service Institute, 1950.

(Received May 6, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 21–29

# EFFECT OF STIMULUS AND INSTRUCTIONAL VARIABLES IN AN AMBIGUOUS CONCEPT–ATTAINMENT TASK[1]

MARY JANKE[2]

*Stockton State Hospital, Stockton, California*

One hundred and seventy-six *S*s categorized, with probabilistic feedback, dots (relevant dimension spatial position), numbers (relevant dimension numerical magnitude), dots with four bits of irrelevant information, and numbers with added position information (redundancy in the relevant variable). The *S*s performed better ($p < .001$) when given a clue to optimal solution. Number categorization was inferior to dot categorization ($p < .01$); irrelevant information did not degrade dot categorization significantly; but redundancy improved number categorization ($p < .05$) when no clue was given to solution. Inferior performance with numbers seems due not to irrelevant information but to lack of salience of the relevant dimension coupled with difficulty in finding the center of the number range. Redundant position information probably mitigated both of these factors.

An ambiguous concept-attainment (CA) task may be defined as one in which the stimuli to be categorized present no cues perfectly correlated with their class membership. Thus the feedback from *E* to *S* is probabilistic rather than deterministic, and *S* can never expect to achieve errorless performance. Two types of probabilistic feedback are misinformative feedback (MF), in which *E* "lies" to *S* on a certain percentage of trials, and noncontingent feedback (NCF), in which feedback is completely unrelated to *S*'s responses (Rogers & Haygood, 1968). Another type of probabilistic feedback is introduced through the use of externally distributed stimuli (Lee, 1963), which are analogous to the (internally distributed) noise and signal plus noise of a signal detection task (Swets, Tanner, & Birdsall, 1961). Here the relevant dimension is composed of an ordered set of stimulus values. In a two-category ambiguous CA task, two overlapping subsets of these values are defined as the categories to be identified; within each subset, the number of stimuli presented which have a particular value is a Gaussian function of that value. Inside the area of overlap of the two subsets, feedback is probabilistic. The *S*'s task may be likened to that of identifying the sex of persons when only their heights are given—distributions of heights are roughly normal for each sex and the distributions overlap, so that while heights at either extreme may be linked to sex with a high degree of confidence, in the middle of the scale there is a range of ambiguity. Although, as mentioned, errorless performance is for all practical purposes impossible in this sort of task, accuracy increases as the separation between means of the distributions increases (Lee & Zentall, 1966). And, though not necessarily correct as defined by feedback, all responses can be optimal in the sense that they maximize the probability of being correct.

The ambiguity present in the CA task of this study was generated by the use of externally distributed stimuli. It has been found that type of stimulus used is a relevant factor in such a task; *S*s categorizing dots varying only in their position along the long axis of a card perform more optimally than do *S*s categorizing two-digit numbers varying in magnitude (Lee & Janke, 1964; 1965). It seemed that one reason for this might be the greater amount of irrelevant information contained in numbers as compared to dots. Since amount of irrelevant information has been

---

[1] The author wishes to thank Natalie Coleman and Terry Maple for recruiting and running *S*s. The opinions or conclusions stated in this article are those of the author and are not to be construed as official or as necessarily reflecting the policy of the California Department of Mental Hygiene.

[2] Requests for reprints should be sent to Mary Janke, Staff Psychologist, Stockton State Hospital, 510 East Magnolia Street, Stockton, California 95202.

TABLE 1

Specifications of Distributions 1 and 2

| Statistic | Relevant dimension | | | |
|---|---|---|---|---|
| | Dot position | | No. magnitude | |
| | 1 | 2 | 1 | 2 |
| *M* | 3.31 | 4.56 | 37.51 | 45.49 |
| *SD* | 1.00 | 1.00 | 6.40 | 6.40 |

Note.—$N$ = 100.

shown to be a significant variable in both deterministic and probabilistic CA experiments (e.g., Archer, Bourne, & Brown, 1955; Johannsen, 1962; Pishkin, 1960; Walker & Bourne, 1961), four bits of irrelevant information were added to dot stimuli to see if this would degrade performance reliably, and to see how the performance level would compare with that achieved with numbers (Janke, 1968). (The variety of irrelevant cues in number stimuli appears equivalent to more than four bits of information. These cues include oddness or evenness of either digit of the number, characteristics of the sum, product, difference, or quotient of the digits, the second digit's being less than five or not, characteristics of the configuration of the digits—straight line vs. curved, e.g.—and so forth.) Although categorization of these "irrelevant dots" tended to be somewhat inferior to categorization of dots, the trend was not significant. In the same study, the hypothesis was tested that presentation of numbers in a manner equivalent to presentation of dots (i.e., in varying positions along the long axis of a card) would improve the performance of *S*s categorizing those numbers. Such presentation, assuming that magnitude corresponded to position, would introduce redundancy in the relevant variable, and redundancy has been shown to be helpful in deterministic CA (Bourne & Haygood, 1959; Keele & Archer, 1967). Again, the trend of the results, while consistent with this hypothesis, was not significant. Because of the relatively small number of *S*s used in this study (10 in each cell of a 4 × 2 factorial design), and the great variability among *S*s, it was decided to increase the size of the groups, still with the goal of explicating the difference in performance levels between probabilistic categorizing of dots and of numbers.

Janke (1968) found a significant superiority for performance under Cue instructions, in which *S*s were told the relative positions of the two categories to be identified, as compared to performance under Neutral instructions, in which no aids to CA were given. This finding accounted for different average performance levels found by Lee and Janke (1964), who used Cue instructions, and Lee and Zentall (1966), who used Neutral instructions. The instructional variable was also included in the present study. It might be expected that the effect of irrelevant or redundant relevant information would be exaggerated under Neutral as compared with Cue instructions, producing an interaction between instructions and stimuli. To put it another way, if *S*s are in essence told the optimal solution of the CA task, the presence of irrelevancy or redundancy in the stimuli might be expected not to have so great an effect as otherwise.

## METHOD

*Stimuli.*—Stimuli were those used by Janke (1968). Two overlapping normal distributions were specified on a continuum ranging from 1 through 42 for dot stimuli (21 through 62 for number stimuli, in order to have two digits in each number). Nominal means, standard deviations, and frequencies of these distributions are given in Table 1; means and standard deviations are in inches from the left side of the stimulus card for the dot-position continuum and in the number continuum units for the number continuum.

Each stimulus value, or sample point, represented .01 of the area of the distribution from which it derived. Sample points were placed at the medians of their intervals, with the lowest point of Distribution 1, for example, located at the point whose normal deviate corresponds to a left-tail area of .005. The value for this sample point was found by solving the equation $M_1 + z\sigma_1 = r$ for the real number $r$ corresponding to $z$ on the stimulus continuum and rounding $r$ to an integer $n$ to give the empirical stimulus value. $\sigma_1$ (as well as $\sigma_2$) was set at 1.00 and $M_1$ was chosen so that $n$ would equal 1, subject to the restrictions that $(M_2 - M_1)/\sigma = 1.25$ and that the highest point in Distribution 2 would be 42. ($(M_2 - M_1)/\sigma$, or $d'$, has been used in signal detection theory as an index of discriminability.)

Other sample points were found by adding successive increments of .01 to the original .005 and using the same equation to find new *r*s and *n*s; a similar procedure was used to find the sample points for Distribution 2. Calculations were performed by computer. The lowest point in Distribution 1 was 1; the highest point was 34. The lowest point in Distribution 2 was 9, and the highest point was 42. The point at which the two distributions crossed had Stimulus Position 21.5. There were no stimuli actually occupying this position, as presented stimulus positions always had integral values. To the left of (below) this position a stimulus sample was more likely to be a 1 than a 2; to its right (above) the reverse was true.

After having been specified, the sample points were put on cards, one to a card. Four decks of 400 cards each were constructed; each deck contained the 200 sample points from Distributions 1 and 2 in random order followed by the same 200 sample points in a different random order. These orders were the same for all four decks. The distribution from which each sample point derived was written on the back of its card.

The four decks represented the four types of stimuli used in the experiment—dots, irrelevant dots, numbers, and redundant numbers. A card for dot stimuli (D) was a 5 × 8 in. plain file card with an inked circle about $\frac{1}{8}$ in. in diameter placed at one of 42 equally spaced positions along an imaginary horizontal line 2.5 in. from the upper, long edge of the card. The dot corresponding to Stimulus Position 1 was 1 in. from the left edge of the card; the dot corresponding to Stimulus Position 42 was the same distance from the right edge, so that the point at which the distributions crossed was located at the card's midline.

A card for irrelevant dot stimuli (ID) was the same as a card for dot stimuli, except that the dots varied along four binary irrelevant dimensions, as well as in horizontal position on the card. These dimensions were color (red vs. blue), size ($\frac{1}{8}$ in. diameter vs. $\frac{1}{4}$ in.), interior (filled vs. open), and vertical position (3.1 in. from the top of the card vs. 1.4 in.). The values of the irrelevant binary variables were assigned randomly to the sample points.

A card for number stimuli (N) was a 3 × 5 in. plain file card with an integer between 21 and 62, inclusive, printed in ink at its center. Digits were drawn uniformly by means of a lettering set, and were .33 in. high.

A card for redundant number stimuli (RN) was a 5 × 8 in. plain file card with an integer between 21 and 62, inclusive, printed in ink at the position on the card which a dot with the equivalent stimulus position would occupy in the deck of dot stimuli. This yielded a perfect correlation between number magnitude and the distance of the number from either vertical edge of the card. Digits were drawn uniformly by means of a lettering set and were .19 in. high.

*Subjects and design.*—The *S*s run in 1970 and 1971 were 80 undergraduate men and women from the University of the Pacific and 40 undergraduate men and women from the University of California at Davis. In order to bring the total to 176 (22 *S*s in each cell of the design), 56 answer sheets were selected, randomly within cells and distributed fairly evenly across conditions, from Janke (1968). These answer sheets gave the responses of undergraduate men and women at Sacramento State College who had served as *S*s in 1967. As mentioned, the same stimulus materials were used for all of these *S*s and procedure and instructions were identical, with the exception that *S*s run in 1967 were paid $1.50 for their participation and later *S*s were unpaid. Each *S* of the 176 served in one condition of a 4 × 2 factorial design. These factors were stimulus type (four levels) and instructions (two levels).

*Procedure and instructions.*—The *S*s were run in pairs seated side by side at a table upon which, about 3 ft. away, a card rack was displayed. They were prevented from seeing one another's answers by means of a screen. The *E* stood behind the table to place cards one at a time in the card rack; another screen hid the pile of cards from *S*s until the time they were displayed. The order in which experimental conditions were to be run was determined randomly prior to the experiment and *S*s were assigned to these conditions in the order of their appearance at the laboratory.

Each *S* was given two dittoed answer sheets, stapled together, with spaces for 400 answers. Then either Cue or Neutral instructions were read aloud by *E*. Neutral instructions told *S*s that each card was either a 1 or a 2 and that they should write 1 or 2 in the corresponding space on the answer sheet, without hesitating, whenever they were shown a card. They were told that although at first they would be guessing, they would learn about the cards with experience. They were also told not to worry about getting every answer right, just to try to do as well as they could on the average. Cue instructions gave this information plus the fact that, on the average, cards that were 2s had dots farther to the right (higher numbers on them) than did cards that were 1s. Instructions to RN Cue *S*s mentioned number magnitude only, not spatial position. After *S*s had indicated that they understood the instructions, the first card was shown in the card rack; they wrote their answers and were then told by *E* whether the card had been a 1 or a 2. From then on, the procedure was paced by *S*s, through the deck of 400 cards. This always took less than an hour.

After all the cards had been shown, *S*s run in 1970 and 1971 were asked what hypothesis they had last been using; RN *S*s were asked which they had paid most attention to—the numbers themselves or their position on the card—and N and RN *S*s in the Cue instruction condition were asked what the average number was. The *S*s run in 1967 were asked only how the 1s and 2s were arranged along the stimulus-position continuum.

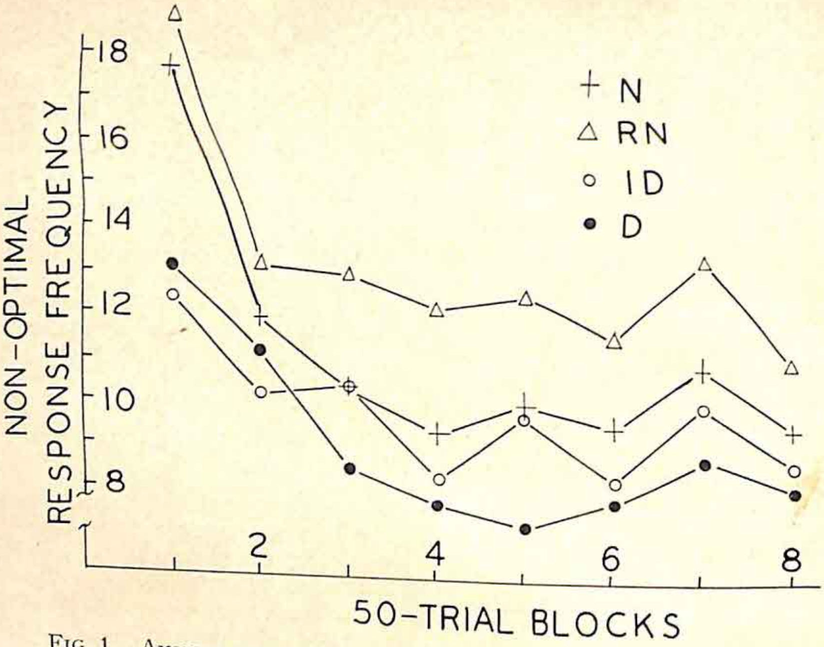


FIG. 1. Average nonoptimal response frequency for stimulus groups under Cue instructions, blocks of 50 trials.

## RESULTS

*Nonoptimal response frequency.*—The optimal response frequency is the number of responses which maximize *S*'s chance of being "correct"—i.e., of giving an answer which agrees with *E*'s feedback. In general, for this task, it is the number of responses "1" to stimuli falling to the left of the point at which the two stimulus-position distributions intersect, plus the number of responses "2" to stimuli falling to the right of this point. Nonoptimal response frequency is the reverse—the number of responses which minimize *S*'s chance of "correctness" by assigning a 1 to stimuli at the right of the intersection point or a 2 to stimuli at the left. It was used as the response measure in this study for convenience in calculation, and may be likened to a measure of errors.

Nonoptimal response frequency is for this study the number of responses "2" to stimuli having a stimulus position of 21 or less (41 for numbers) and of responses "1" to all other stimuli. Nonoptimal response frequencies were found for blocks of 50 trials and means for the several stimulus conditions were plotted for Cue (Fig. 1) and Neutral (Fig. 2) groups; nonoptimal response totals for the last 200 trials were analyzed. Previous studies (e.g., Janke, 1968) have shown that by the second block of 200 trials, *S*s have reached relative stability of response, and analysis was restricted to these data for the purpose of comparison with other studies of asymptotic response in ambiguous CA tasks using externally distributed stimuli (Lee & Janke, 1964; 1965; Lee & Zentall, 1966). Means and standard deviations of nonoptimal response frequencies over the last 200 trials for stimulus groups as a function of

TABLE 2

MEANS AND STANDARD DEVIATIONS OF NONOPTIMAL RESPONSE SCORES, LAST TWO HUNDRED TRIALS

| Stimuli | Instruction | | | |
|---|---|---|---|---|
| | Cue | | Neutral | |
| | *M* | *SD* | *M* | *SD* |
| D | 31.9 | 15.5 | 59.8 | 25.2 |
| ID | 36.8 | 21.9 | 71.5 | 25.8 |
| RN | 47.7 | 22.7 | 69.1 | 30.8 |
| N | 39.4 | 17.4 | 86.5 | 22.7 |

Note.—Abbreviations: D = dot stimuli, ID = irrelevant dot stimuli, RN = redundant number stimuli, N = number stimuli.

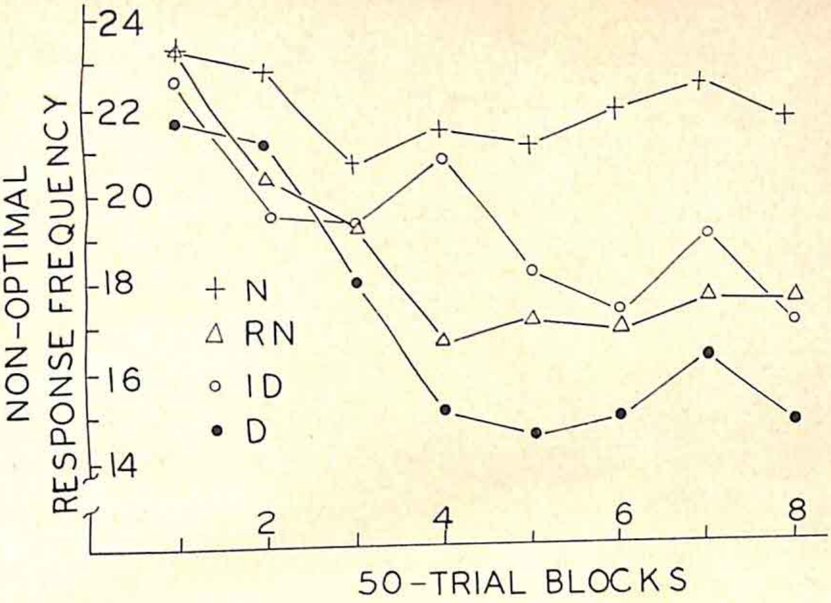


FIG. 2. Average nonoptimal response frequency for stimulus groups under Neutral instructions, blocks of 50 trials.

instructions are shown in Table 2. If *S*s had probability matched at each stimulus position (probability micromatching; Lee, 1963), their score would have been 54 (27% nonoptimal responses).

In order to proceed with the analysis, it was first necessary to determine whether the three populations of *S*s, run at different times by different *E*s, were sufficiently comparable for their results to be combined. An analysis of variance for unequal subclass frequencies, as described by Walker and Lev (1953, p. 381), was performed for nonoptimal response frequencies over the last 200 trials. This analysis was in the form of a $3 \times 8$ factorial design, with 3 levels of *S* population and 8 levels of stimulus-instruction condition. Only the main effect of conditions was significant, $F$ (7, 152) = 13.39, $p < .001$, and there was no significant interaction. Variances as well as means of the 3 *S* populations were comparable, $F_{\text{max}} = 1.09$, $k = 3$, and $n = 79$, the larger of the two *n*s.

Variances of the eight stimulus-instruction groups were also not significantly different, $F_{\text{max}} = 3.93$, $k = 8$, $n = 21$. It was concluded that analysis of variance could be performed on the combined data. This analysis showed a significant effect of instructions, $F$ (1, 168) = 87.87, $p < .001$, and of stimuli, $F$ (3, 168) = 4.44, $p < .01$. The interaction approached significance, $F$ (3, 168) = 2.40, $p < .10$. Cue instructions (39.0 or 19.5%) were associated with better performance than were Neutral instructions (71.7 or 35.8%). With respect to stimulus effects, Newman-Keuls tests for differences between treatment means, with $MS_e = 535.98$, revealed that Group D (45.8 or 22.9%) was superior both to N (63.2 or 31.6%), $p < .01$, and to RN (58.4 or 29.2%), $p < .05$. Groups N, RN, and ID did not differ significantly.

Since the $F$ for interaction between stimuli and instructions was above the .10 level, it was decided to test for differences among stimulus groups at each instruction level, using the Newman-Keuls procedure. Cue groups did not differ significantly among themselves. For Neutral groups, N (86.5 or 43.2%) was significantly inferior both to RN (69.1 or 34.6%), $p < .05$, and to D (59.8 or 29.9%), $p < .01$.

*Estimates of average number: Cue groups.*—The *S*s in N and RN groups run under Cue instructions in 1970 and 1971 were asked at the end of the experimental session to estimate the average number in the set of numbers shown to them. Some

*S*s run in 1967 offered estimates of the average number also, so center estimates were received from 18 *S*s in each group, or a total of 36 *S*s. The actual average of presented stimuli was 41.5; the mean estimate of *S*s in Group N was 38.5 and that of *S*s in Group RN was 40.0. The difference in average estimates for the two groups was not significant. The variance of center estimates was 12.24 for Group N and 4.44 for Group RN; this difference was significant, $F\ (17, 17) = 2.76, p < .05$.

*Reported attention to redundant information.*—The *S*s in RN groups run in 1970 and 1971 were asked at the end of the experimental session whether they had paid more attention to what the numbers were or to their position on the card (RN Neutral) or whether they had paid attention to where the numbers were on the card and used this information (RN Cue). The forms of the questions were slightly different because *S*s in Neutral groups had been told nothing in the instructions regarding position or number magnitude, while *S*s in Cue groups had been told that numerically lower numbers were usually 1s and numerically higher numbers usually 2s. In the RN Neutral group, 15 *S*s responded. There was an inverse monotonic relationship between their reported use of the position cue and their nonoptimal response frequencies. For *S*s who reported paying attention mostly to where the numbers were on the card ($n = 2$) the average number of nonoptimal responses was 24 (12%); for *S*s who reported using position and number equally ($n = 3$) it was 57 (28.5%); for *S*s who reported paying most attention to what the numbers were ($n = 8$) it was 87.2 (43.6%), and for *S*s who reported paying attention neither to position nor to number, i.e., answering randomly ($n = 2$), it was 92.5 (46.2%). An extension of the median test showed that the relationship between nonoptimal response score and *S*'s report was significant, $\chi^2\ (3) = 7.83, p < .05$. (Of the 8 *S*s who reported that they paid attention to what the numbers were rather than to where they were, only one claimed that a hypothesis based on numerical magnitude was his last-used hypothesis. The *S*s who paid attention to position, on the other hand, were forced to use left *vs.* right as a cue, since this was the only way in which position varied. The 2 *S*s who claimed to have paid attention primarily to position explicitly reported left *vs.* right as their last-used hypothesis, as did 1 of the 3 *S*s who reported paying attention to number and position equally.) In the RN Cue group, 17 *S*s responded; 12 reported using position as a cue, and 5 claimed to have paid no attention to position. The mean nonoptimal response score for the former group was 36.1 (18.0%) and for the latter group it was 70.4 (35.2%). This difference was significant, $t(6) = 2.94, p < .025$.

## DISCUSSION

The results of this study support the conclusion that instructional differences can be an important factor in an ambigous CA task. Performance under Cue instructions is superior to that under Neutral instructions, and there is a suggestion that Cue instructions may be especially helpful to *S*s categorizing the most complex stimuli (nonredundant numbers). This is consistent with Haygood and Bourne's (1965) finding that *S*s required to learn both the rule and the set of relevant attributes in a deterministic CA task were handicapped with respect to *S*s told either of these at the outset. The Cue instructions used here informed *S*s both of the relevant attribute and of a rule to guide their assignment of responses to values of that attribute.

It may seem surprising that this instructional effect persists through the last 200 trials, as one would perhaps expect that all *S*s, whether cued or not, would have discovered the relevancy of stimulus position by this stage of the task. It is perhaps true that Neutral *S*s hypothesize a relationship between stimulus position and response category; but, if so, they apparently tend, more than do Cue *S*s, to discard the hypothesis as having too many exceptions to be correct. Cue *S*s, on the other hand, have been given this hypothesis in the form of a rule by *E*, and therefore may be considerably less prone to discard it. Even these latter *S*s, however, seem to modify the rule through the use of supplementary (irrelevant) cues instead of relying exclusively on it. No *S* gave zero nonoptimal responses, and only two closely approached this level of performance.

Stimulus factors made a difference here as in earlier studies (e.g., Lee & Janke, 1965); categorization of dots remained superior to categorization of numbers. The quantity of irrelevant information contained in number stimuli was probably not the critical factor in the inferior performance of number-categorizing *S*s under the conditions of this experiment, although it may have contributed to the difficulty of the task. This is inferred from the fact that four bits of irrelevant information did not significantly degrade the performance of *S*s categorizing dots. This finding is puzzling. Although two recent CA studies (Byers & Davidson, 1968; Dominowski, 1969) have reported a decremental trend but no significant effect of added irrelevant information, the bulk of the evidence (e.g., Walker & Bourne, 1961) shows a linear decrement in CA with increases in irrelevant stimulus information. It seems possible, however, that under some experimental conditions the effect of quantity of irrelevant information levels off after a critical value has been reached, so that for some large $n$ and relatively small $k$ the effect of $n + k$ bits of irrelevant information is not much greater than the effect of $n$ bits. Such conditions might obtain in a situation where *S*s have more irrelevant information to process than they can deal with in the time allowed, so that they cannot in general reach full attainment of the concept. This may have been the case in the present study in that, first, the probabilistic feedback insured that no errorless solution would be found. This led *S*s, as they reported, to resort to the use of irrelevant reinforcement-sequence cues and to random guessing, presumably determined by some set of irrelevant internal cues. Such irrelevant cues were available to all *S*s, including those in Cond. D, which ostensibly involved no presentation of irrelevant information. Second, the experiment terminated after a fixed number of trials, and at the point of termination no *S* had stabilized at 100% optimal responses, although 2 of the 176 *S*s gave 95% or more optimal responses over the last 200 trials. These conditions may account for the finding that the four added bits of irrelevant information in Cond. ID were not significantly detrimental to performance.

A different explanation for the equivalence of D and ID might be that restriction of the analysis to the last 200 trials is responsible, in that the four irrelevant dimensions may have been tested and discarded before this point. However, one would then expect the difference between ID and D in the first 200 trials to have been greater than in the last 200, perhaps reaching significance. This was not the case: ID and D scores were closer together in the early trials than in the later ones for both Cue and Neutral groups (Fig. 1, 2). Moreover, in both instruction groups, ID *S*s' reports of their last-used hypotheses commonly included mention of irrelevant stimulus cues, so use of these cues apparently had not been abandoned.

Still another hypothesis might be that binary irrelevant dimensions lack the salience of quasi-continuous dimensions, thus providing little distraction for *S*s. This is entirely possible and deserves study, although *S*s' reports mentioned above indicate that for many *S*s (at least 40%) the irrelevant cues were still in use at the end of the experimental session.

In the present study, there was no reliable difference among stimulus groups in the Cue condition. Lee and Janke (1964), on the other hand, found significantly inferior performance with numbers as compared to dots under Cue instructions. There were other differences in results between the two experiments; dot-categorizing *S*s in the 1964 study performed close to the optimal level (about 6% "crossover choices," a measure similar to nonoptimal responses), while in the present study, D Cue *S*s gave, on the average, 16% nonoptimal responses. The N Cue *S*s in the present study gave 20% nonoptimal responses; number-categorizing *S*s in the Lee and Janke (1964) study gave an average of around 15% crossover choices. These differences, both in overall performance level and in the dot-number discrepancy, may be due in part to the fact that the two stimulus-position distributions used by Lee and Janke (1964) did not overlap as extensively as did the distributions used here. The discriminability index, $d' = (M_2 - M_1)/\sigma$, was 1.50 for the earlier study; here it was 1.25. This means that *S*s, in following the rule given by the Cue instructions, would make fewer apparent errors (agree more frequently with *E*'s feedback) under the conditions of the earlier study. It seems likely that *S*s, having been told this rule, still search for useful supplementary cues (stimulus cues for number-categorizing *S*s, reinforcement sequence and internal cues for all *S*s) in trying to reach what is to them a satisfactory level of performance. The fewer apparent errors they are making in following the rule, the shorter this search may be. A shorter search should improve performance generally since the supplementary cues are in fact irrelevant; it may also improve the performance of dot-categorizing *S*s to a greater extent than that of number-categorizing *S*s, if the length of the abbreviated search remains proportional to the number of irrelevant cues available and if in addition the number of trials or time allowed is great enough for the performance-degrading search for a useful supplementary cue to be generally abandoned by dot-categorizing *S*s, but not by number-categorizing ones. This may have happened in the Lee and Janke (1964) study. Thus, depending upon such factors as the overlap of the two stimulus-position distributions, quantity of irrelevant information in an ambiguous CA task may be hypothesized to have a greater or lesser effect on performance. This hypothesis seems to warrant further study.

A clue to the cause of the difference in the present study between performance with dot stimuli and performance with number stimuli is offered by the finding that under Neutral instructions, redundant position information improved the categorization of numbers. This finding agrees with evidence from deterministic CA studies (e.g., Bourne & Haygood, 1959). For optimal performance in the number categorization task, it was necessary first to hit upon the idea of dividing the range of numbers at some central number and second to determine what this central number was. The first requirement was somewhat abstract; it became more concrete when stimuli could be seen varying in their actual physical positions from right (of center) to left on a card. This added concreteness, leading to increased salience of the relevant dimension, may have facilitated *S*s' meeting of the first requirement, either through attracting their attention to the dimension of numerical magnitude or through leading them to ignore the numbers as numbers, considering them merely as marks varying in position. The center of the card, moreover, offered a natural dividing point, probably facilitating the meeting of the second requirement. Some support for these speculations comes from the findings that RN *S*s reporting use of position as a cue had fewer nonoptimal responses than those who did not, that RN Neutral *S*s who used number rather than position as a cue did not in general use numerical magnitude as the basis for their final hypotheses, and that in Cue groups (which were the only ones asked to estimate average number magnitude), variance of these estimates was considerably larger for *S*s categorizing numbers than for *S*s categorizing redundant numbers. It is not easy to estimate the central tendency of a set of numbers presented singly in random order. Lee and Janke (1964) compared their dot-categorizing *S*s with *S*s who were shown the usual number stimuli and instructed merely to respond "1" if the stimulus was below the median of all stimulus samples seen, and "2" if above. Dot-categorizing *S*s were still significantly superior, as measured by percentage crossover choice.

The finding that relevant redundancy had no facilitating effect in the Cue condition is of some interest. There is no significant difference between N Cue and RN Cue groups, and in fact the trend over the entire 400 trials is toward inferiority of RN Cue *S*s (Fig. 1). One would expect that RN Cue *S*s would be somewhat superior to N Cue *S*s, due to facilitation in finding the center of the number range. Since all Cue *S*s knew the relevant dimension from the outset, this effect would probably show itself most clearly during the earliest blocks of trials, and Fig. 1 does indicate that nonoptimal response frequency levels off more rapidly in Group RN Cue than in Group N Cue, which continues a marked improvement over the first 200 trials. The lack of superiority of Group RN Cue, even in the earliest trials, may have stemmed from the operation of an opposing factor, such as confusion or added impetus to hypothesis testing in the presence of another stimulus dimension. But this seems unlikely in view of the superiority of RN Neutral *S*s over N Neutral *S*s and the lack of a reliable effect of four added irrelvant dimensions on dot categorizing.

In conclusion, it seems that a lack of salience of the number-magnitude dimension and, secondarily, difficulty in determining the center of the number range are the major factors accounting for the inferior performance of *S*s trying, under the conditions of the present study, to attain an ambiguously defined concept based on numerical magnitude.

## REFERENCES

Archer, E. J., Bourne, L. E. Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology*, 1955, **49**, 153–164.

Bourne, L. E. Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification. *Journal of Experimental Psychology*, 1959, **58**, 232–238.

Byers, J. L., & Davidson, R. E. Relevant and irrelevant information in concept attainment. *Journal of Experimental Psychology*, 1968, **76**, 277–281.

Dominowski, R. L. Concept attainment as a function of instance contiguity and number of irrelevant dimensions. *Journal of Experimental Psychology*, 1969, **82**, 573–574.

HAYGOOD, R. C., & BOURNE, L. E. JR. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, **72**, 175–195.

JANKE, M. *Probabilistic categorizing behavior: Stimulus and instructional factors.* (Doctoral dissertation, University of California, Berkeley) Ann Arbor, Mich.: University Microfilms, 1968. No. 68–13, 914.

JOHANNSEN, W. J. Concept identification under misinformative and subsequent informative feedback conditions. *Journal of Experimental Psychology*, 1962, **64**, 631–635.

KEELE, S. W., & ARCHER, E. J. A comparison of two types of information in concept identification. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 185–192.

LEE, W. Choosing among confusably distributed stimuli with specified likelihood ratios. *Perceptual and Motor Skills*, 1963, **16**, 445–467.

LEE, W., & JANKE, M. Categorizing externally distributed stimulus samples for three continua. *Journal of Experimental Psychology*, 1964, **68**, 376–382.

LEE, W., & JANKE, M. Categorizing externally distributed stimulus samples for unequal molar probabilities. *Psychological Reports*, 1965, **17**, 79–90.

LEE, W., & ZENTALL, T. R. Factorial effects in the categorization of externally distributed stimulus samples. *Perception & Psychophysics*, 1966, **1**, 120–124.

PISHKIN, V. Effects of probability of misinformative feedback and number of irrelevant dimensions upon concept identification. *Journal of Experimental Psychology*, 1960, **59**, 371–378.

ROGERS, S. P., & HAYGOOD, R. C. Hypothesis behavior in a concept-learning task with probabilistic feedback. *Journal of Experimental Psychology*, 1968, **76**, 160–165.

SWETS, J. A., TANNER, W. P. JR., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301–340.

WALKER, C. M., & BOURNE, L. E. JR. The identification of concepts as a function of amounts of relevant and irrelevant information. *American Journal of Psychology*, 1961, **74**, 410–417.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Henry Holt, 1953.

(Received May 12, 1971)

*Journal of Experimental Psychology*
*1972, Vol. 93, No. 1, 30–34*

# BINOCULAR RELATIONSHIPS IN A SIZE AND COLOR ORIENTATION SPECIFIC AFTEREFFECT[1]

GERALD M. MURCH[2]

*Portland State University*

The McCollough Effect, an orientation-specific color effect, was established monocularly using the aftereffect inspection pattern of C. Blakemore and P. Sutton oriented vertically on a red ground, which alternated with the inspection pattern oriented horizontally on a green ground. Testing for the size aftereffect and the McCollough Effect was carried out with neutral vertically and horizontally oriented test patterns presented either to the same eye used during inspection or presented to the opposite eye. The size aftereffect showed interocular transfer as it occurred in both conditions while the McCollough Effect was observed only in the same eye test condition. The results are interpreted to mean that the McCollough Effect involves color adaptation at a visual locus separate from orientation-specific receptors.

The McCollough Effect, an orientation-specific color aftereffect, first observed by Celeste McCollough (1965), consists of the apparent greenish coloration of a test pattern of horizontal lines and a reddish coloration of a test pattern of vertical lines after *O* has watched inspection patterns of vertical lines on a green background alternate several times with a pattern of horizontal lines on a red background.

Taking her lead from the early work of Hubel and Wiesel (1959, 1962), in which orientation-specific edge detectors were found in the striate cortex of the cat, McCollough (1965) reasoned that her orientation-specific aftereffect might be due to the chromatic adaptation of vertical and horizontal edge detectors. Specifically, she assumed that edge detectors were both color coded and sensitive to fatigue after prolonged stimulation. The repetitive presentation of a pattern of red, horizontal lines would render the horizontal detectors insensitive to red such that a subsequent test pattern of horizontal lines on white would be processed by the unadapted detectors. Thus the pattern would appear greenish. Conversely, the prolonged presentation of green vertical should serve to fatigue the green vertical detectors and cause the neutral test pattern of vertical lines to appear tinged with red or pink.

Several other explanations of the McCollough Effect have been offered. Harris and Gibson (1968) developed a hypothetical sensory unit, the dipole, which responded to differences in the intensity of stimulation to the respective poles of the unit. Stromeyer (1969) has elaborated upon the model of McCollough and assumes that the color-coded edge detectors are organized as opponent process units sensitive to vertical and horizontal edges as well as color pairs (e.g., green/red). Characteristic of these models is the assumption that the adaptation for color and/or line orientation occurs within a single sensory unit. However, more recent research by Hubel and Wiesel (1968) reveals that color-sensitive elements and edge detectors may be located in two distinct anatomical areas of the visual system. In their study of the primate, they located a preponderance of opponent process color cells in the lateral geniculate nucleus (LGN), while in the cortex most units observed were sensitive to line orientation but not to color.

These observations led Murch and Hirsch (in press) to propose that the McCollough Effect may well be the result of color adaptation *in conjunction with* a specific line orientation rather than as a result of

---

[1] This research was supported by Research Grant NSF GB 11817 from the National Science Foundation, Division of Psychobiology. The author wishes to express his thanks to Barbara Breck for her aid in data collection for the experiment reported and Joy Hirsch for her aid in data analysis.

[2] Requests for reprints should be sent to Gerald M. Murch, Department of Psychology, Portland State University, Portland, Oregon 97207.

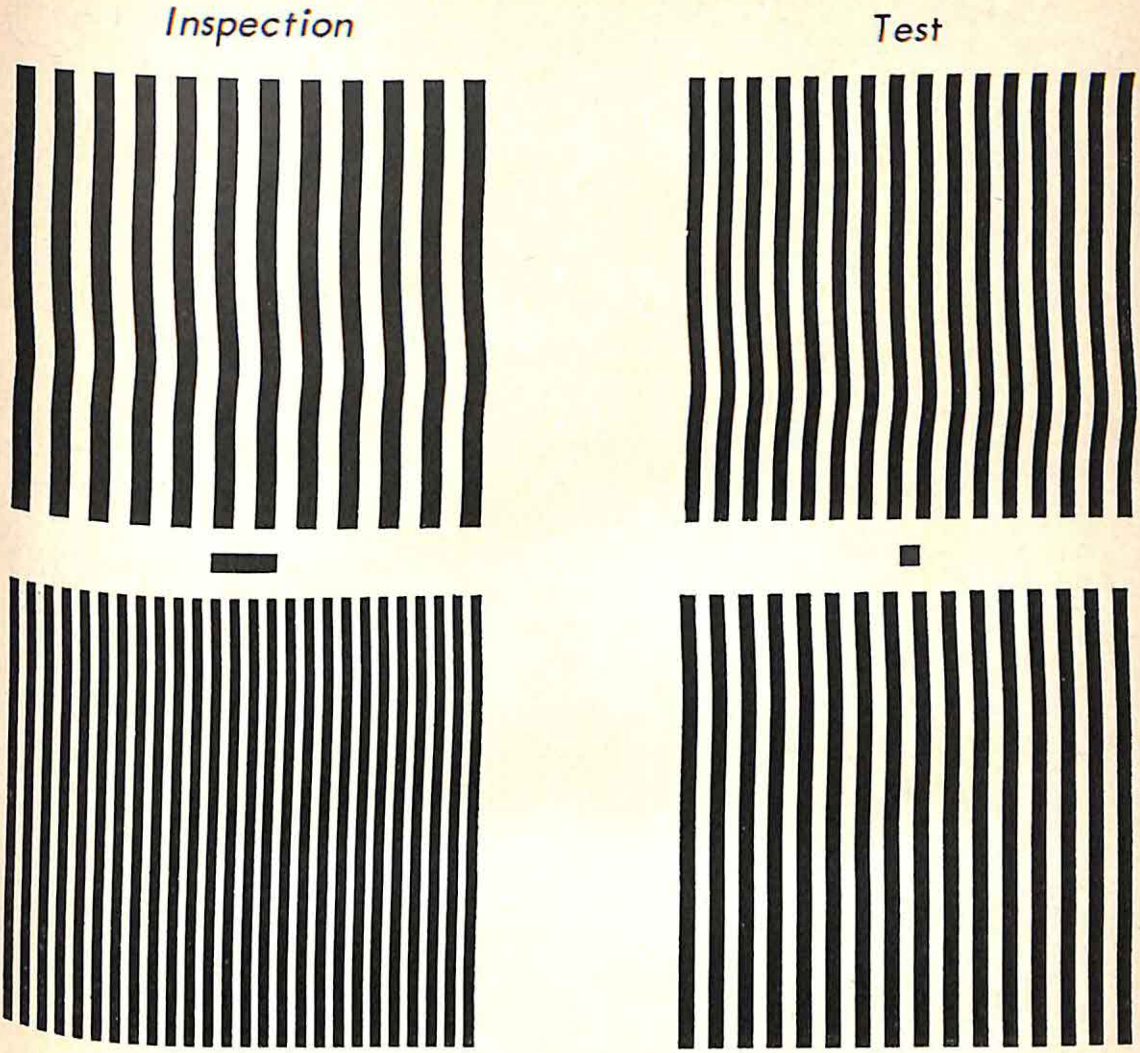


FIG. 1. Inspection and test patterns (after Blakemore & Sutton, 1969).

adaptation of color-coded edge detectors as proposed by McCollough (1965) and Stromeyer (1969).

The present experiment sought to test the possibility that the color adaptation in the McCollough Effect does not take place in the edge detectors. A technique for such a test suggests itself from the research of Blakemore and Sutton (1969), who have reported a new type of size aftereffect. The *S*s viewed a pattern of wide lines on one-half of a fixation area with narrow lines on the other half. A test figure of lines of equal density on the left and right was reported to appear narrower on the side previously stimulated by the wide inspection lines while the area initially stimulated by the narrow lines of the inspection grid appeared wider on the test pattern. Because this aftereffect can involve either vertically or horizontally oriented grids, it might serve as the lined pattern for a McCollough Effect.

By alternately presenting a vertically oriented inspection figure of half wide and half narrow lines on a red background with a horizontally oriented figure of wide and narrow lines on a green background, one could test for the appearance of both the figural aftereffect and the McCollough Effect. Through the presentation of the inspection patterns to one eye and the test patterns to the other, unstimulated eye, one could test for the interocular transfer

TABLE 1

*The Ss Reporting the Size Aftereffect and McCollough Effect under Condition 1 and Condition 2*

| Test group | *N* | Percentage |
|---|---|---|
| *S*s reporting size aftereffect | | |
| Cond. 1 | | |
| Vertical | 20 | 83 |
| Horizontal | 20 | 83 |
| Cond. 2 | | |
| Vertical | 21 | 87 |
| Horizontal | 17 | 71 |
| *S*s reporting McCollough effect | | |
| Cond. 1 | | |
| Vertical | 14 | 58 |
| Horizontal | 11 | 46 |
| Cond. 2 | | |
| Vertical | 1 | 4 |
| Horizontal | 1 | 4 |

Note.—Cond. 1: inspection and test to same eye, $n = 24$; Cond. 2: inspection and test to different eyes, $n = 24$.

of both the figural aftereffect and the McCollough Effect. If the McCollough Effect is mediated by color-coded edge detectors, one would expect either a transfer of both effects or of neither, depending upon whether or not these units are binocularly or monocularly driven. The transfer of one or the other effect would necessarily imply that the adaptation for edges and for color takes place in separate portions of the visual system.

## Method

*Apparatus.*—Two 35-mm. slides were prepared of the inspection figures used by Blakemore and Sutton (1969); these are shown in Fig. 1, as well as two slides of their test figures. The inspection figures were projected via Kodak Carousel slide projectors. One figure was projected in the vertical orientation and passed through a Wratten (#25A) red filter, while the second inspection pattern was horizontally oriented and passed through a Wratten (#61) green filter. Neutral density filters were placed with the chromatic filters and the luminance equated at 10.0 ftc. at the viewing screen. The two achromatic test patterns were also projected, one oriented vertically and the other oriented horizontally. Their luminance levels were matched at 15.0 ftc. These values were obtained by placing the photosensor of a photometer (International Light Model IL 600) in the center of each pattern facing the projector.

All stimulus patterns subtended 8.7° × 17.5° of visual angle at the eye of *S* seated 3 m. from the white cardboard projection screen. The four projectors were located on a pair of shelves above and behind *S*.

For the inspection patterns, the wide lines subtended 22′ of arc while narrow lines each subtended 10′ of arc. For the wide portion this represented a line density of 1.37 lines per degree of visual angle and 2.97 lines per degree for the narrow portion. The two test patterns shared the same overall dimensions with a line width of 16′ of arc and a line density of 1.83 lines per degree of visual angle.

*Subjects and procedure.*—The *S*s were 48 student volunteers from a psychology of perception course. None of the *S*s were familiar with either the size aftereffect or the McCollough Effect. Prior to participation in the experiment, all *S*s were tested via the Devorine plates for color weakness and given ½ hr. of practice matching filters with the projection colorimeter described in detail by Murch (in press).

The *S*s were randomly divided into two groups. For Group 1 both the inspection patterns and the test patterns were presented to the left eye. During the entire experiment *S* wore an eye patch over the right eye. For the 24 *S*s of Group 2 the inspection patterns were presented to the right eye and the test patterns to the left eye. During inspection an eye patch was worn over the left eye and changed prior to testing to the right eye.

Within each of the major groups, *S*s were further subdivided. For one-half of the *S*s within each group, the inspection stimuli consisted of the vertically oriented inspection figure with the wide lines at the top on a red background. This was alternated with the same inspection figure horizontally oriented on a green background with the wide line to the left. Each figure was shown 30 times centered on the projection screen with *S*s instructed to confine their gaze to the fixation area in the center of the figure. The duration of each figure was 15 sec., with a 1-sec. interval between figures. After the last presentation of the inspection figures, *S*s sat in the dark for 5 min. and then were shown either the vertically oriented test pattern or the horizontally oriented pattern. Each *S* was asked to: (*a*) judge which grid of the test pattern appeared more narrow (left versus right or up versus down); (*b*) name any color seen in the white area between the black lines of the test pattern; and (*c*) match any perceived color via the projection colorimeter.

After making these three judgments on the first test pattern, that pattern was turned off and replaced by the other test pattern. All *S*s judged both patterns. Half of the *S*s within each group judged the vertically oriented test pattern first, while the other half responded to the horizontal pattern first.

Also within the two main groups, half were shown the vertically oriented inspection pattern with the wide portion on the bottom and the wide portion of the horizontal pattern on the right. The green and red backgrounds were always the same on the horizontal and vertical figure, respectively.

## RESULTS

The results for the two groups in terms of the number of *S*s reporting the size aftereffect and the McCollough Effect are reported in Table 1. Neither the order nor the orientation (wide lines on top or bottom) affected the judgments of the two test patterns. Therefore in Table 1 the judgments of all *S*s within each group were combined.

While a little more than half of the *S*s reported an interocular transfer of the color aftereffect, only one *S* indicated coloration of the test patterns in the contralateral-eye test condition. This *S* reported that he was very unsure of any perceived color and was in fact guessing. The finding that about half of the *S*s report the McCollough Effect conforms to previous results with the effect (Murch, 1969).

The mean and standard deviation of color matches for vertical and horizontal test patterns in CIE $x,y$ coordinates are reported in Table 2 for the two conditions.[3]

The $x,y$ matches for the same eye inspection and test condition (1) indicate a greenish match for the vertical test pattern and reddish for the horizontal test pattern whereas the matches for Cond. 2 (test and inspection to different eyes) are very similar for each test pattern.

Differences between the matches of the vertical and horizontal test patterns for each *S* were compared via *t* tests for correlated observations (one-tailed) using the resultant vector (vectorial addition of $x$ and $y$) and the tangent ($\sin\theta/\cos\theta$) of the matches. For Cond. 1 in which the inspection and test patterns were presented to the same eye, both the vector, $t\,(20) = 1.989$, $p < .05$, and the tangent, $t\,(20) = 3.325$, $p < .005$, indicated that the colorimetric matches for the vertical and horizontal test patterns were significantly different. For Cond. 2, in which the inspection pattern and test patterns were presented to different eyes, no difference in the matches for vertical and horizontal test pattern were found, vector, $t(20) = -.329$, $p > .10$, and tangent, $t(20) = .088$, $p > .10$.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF COLORIMETER MATCHES IN CIE $x$ AND $y$ COORDINATES OF CHROMATICITY SPACE

| Statistic | Vertical test pattern | | Horizontal test pattern | |
|---|---|---|---|---|
| | $x$ | $y$ | $x$ | $y$ |
| | | Cond. 1 | | |
| $\bar{X}$ | .2223 | .5226 | .2227 | .4904 |
| *SD* | .087 | .138 | .067 | .016 |
| | | Cond. 2 | | |
| $\bar{X}$ | .2225 | .5104 | .2226 | .5100 |
| *SD* | .113 | .127 | .100 | .168 |

## DISCUSSION

The results of the experiment support the contention that the color and line adaptation observed in the McCollough Effect occurs at different points within the visual system. The size aftereffect readily displayed transfer while the McCollough Effect was observed only when the test pattern was presented to the same eye stimulated during inspection.

These observations are commensurate with the physiological data of Hubel and Wiesel (1968) in which color-coded cells were found to predominate in the primate LGN, while orientation-sensitive units were located in the cortex. The LGN color-sensitive receptors were monocularly driven, while the cortical orientation units responded to input from either eye. Because these data parallel the known attributes of the visual system as observed by Hubel and Wiesel (1968), they support the model of the McCollough Effect proposed by Murch and Hirsch (in press) based on these observations. This model holds that the McCollough Effect arises as a result of adaptation of opponent process color receptors located in the LGN which feed into orientation-sensitive units in the cortex. After prolonged viewing of vertical lines on a green ground, adaptation for green occurs for those units feeding into or associated with vertical orientation sensitivity. During subsequent testing a neutral pattern of vertical lines is processed through fatigued color receptors

---

[3] Three *S*s in each experimental group experienced difficulty matching perceived colors with the projection colorimeter. As their matches differed substantially from the other *S*s, they were deleted in the analysis.

*and hence appears reddish. A similar process occurs simultaneously for the red, horizontal* inspection pattern. As the color-sensitive units are monocularly driven, presentation of the test pattern to the unstimulated eye does not produce the McCollough Effect as the fatigued color units are not used in processing the test pattern.

From these results the conclusion seems warranted that the McCollough Effect is neither an afterimage nor an aftereffect of the kind involved in size or figural aftereffects. It is not an afterimage, as the color adaptation does not occur in the retinal color receptors by nature of the fact that the two inspection patterns are presented to the same portion of the retina alternately. Further, the work of Hajos (1968) has shown that the perceived colors of the McCollough Effect are not the same as the colors evoked as an afterimage when the same inspection patterns are used to produce both effects. Finally, Murch (1969) found that the McCollough Effect does not follow Emmert's Law as does the complementary afterimage, and Murch and Hirsch (in press) showed that a complementary afterimage could serve as the inducing stimulus for a McCollough Effect.

The McCollough Effect also fails to fall into the classification of an aftereffect. It shows no binocular transfer. Furthermore, the effect probably does not involve a fatiguing of orientation-sensitive units as produced in the figural aftereffect, but rather, in the McCollough Effect, the orientation-specific units serve to convey the effect due to their processing of a neutral lined test pattern through previously adapted color units.

With regard to the figural aftereffect used here, it is interesting to note that the aftereffect could be established for a vertically and horizontally oriented inspection figure simultaneously. This observation would support the contention that the aftereffect is mediated by the adaptation of two separate and independent orientation-specific units.

## REFERENCES

Blakemore, C., & Sutton, P. Size adaptation: A new aftereffect. *Science*, 1969, **166**, 245–247.

Hajos, A. Psychophysiologische Probleme bei "Farbkonturen" und "Konturfarben." *Studia Psychologica*, 1968, **10**, 254–266.

Harris, C. S., & Gibson, A. R. Is orientation-specific color adaptation in human vision due to edge detectors, or "dipoles"? *Science*, 1968, **162**, 1506–1507.

Hubel, D. H., & Wiesel, T. N. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 1959, **148**, 574–591.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, **160**, 106–123.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture of the monkey striate cortex. *Journal of Physiology*, 1968, **195**, 215–243.

McCollough, C. Color adaptation of edge-detectors in the human visual system. *Science*, 1965, **149**, 1115–1116.

Murch, G. M. Size judgments of McCollough afterimages. *Journal of Experimental Psychology*, 1969, **81**, 44–48.

Murch, G. M. CIE x,y coordinates from an inexpensive projection colorimeter. *Behavior Research Method and Instrumentation*, in press.

Murch, G. M., & Hirsch, J. The McCollough Effect created by complementary afterimages. *American Journal of Psychology* (in press).

Stromeyer, C. F. Further studies of the McCollough Effect. *Perception & Psychophysics*, 1969, **6**, 105–110.

(Received June 4, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 35–42

# MORALITY JUDGMENTS:

## *TESTS OF AN AVERAGING MODEL* [1]

MICHAEL H. BIRNBAUM [2]

*University of California, Los Angeles*

Pairs of items describing objectionable behaviors were rated for their overall morality. Contrary to additive or constant-weight averaging models, the ratings were demonstrated to depend upon the range as well as the average scale value of the component behaviors. A range model accounted for more than half of the variance left unexplained by the additive models. One interpretation of the range effect postulates that each component stimulus produces a distribution of values. The value of the stimulus combination is assumed to be the mean value in the overlap of the component distributions, which is closer to the item with the narrower dispersion.

How immoral is it to both "pocket the tip the previous customer left for the waitress" *and* "poison your neighbor's dog whose barking bothers you?" Anderson (1968b) has suggested that *S*s combine information by averaging psychological values associated with each component stimulus. Thus, *S* independently assesses the morality of pocketing the tip and the morality of poisoning the dog and then averages his two assessments to arrive at an overall rating. According to this view, the psychological value of the *whole* is simply an average of the psychological values of the *parts*. The theory is a special case of a general additive model (Rosenberg, 1968), which can be written:

$$\Psi_{1\ldots k\ldots n} = \sum_{k=0}^{n} w_k s_k + c \qquad [1]$$

where $\Psi_{1\ldots k\ldots n}$ is the psychological value of the combination of $n$ stimuli, $s_k$ is the psychological value of stimulus $k$, $s_o$ and $w_o$ are the scale value and weight of the initial impression, and $c$ is an additive constant. In the usual application of such models, the weights, $w_k$, are assumed to be independent of the scale values, and the effect of each stimulus is assumed to be independent of the other stimuli with which it is presented (Anderson, 1970). The constant-weight averaging model (Anderson, 1968b; Rosenberg, 1968) is a special case of the additive model and requires the additional assumption that the weights sum to unity.

The overt response, $R_{1\ldots k\ldots n}$, is assumed to be a monotonic transformation, $J$, of the psychological impression values:

$$R_{1\ldots k\ldots n} = J(\Psi_{1\ldots k\ldots n}). \qquad [2]$$

Since $J$ is usually assumed to be linear, analysis of variance may be applied directly to the responses obtained in a factorial experiment to test the general additive model (Anderson, 1968b, 1970, 1971). These models predict no interaction between component stimuli, and the test for interactions provides a test of all of the special cases of this model including both additive and constant-weight averaging models.

These models have provided a good fit to the data obtained in previous studies using verbal stimuli (Anderson, 1968b, 1971; Slovic & Lichtenstein, 1971) and have received some support in studies using psychophysical stimuli (Anderson, 1970; Slovic & Lichtenstein, 1971). However,

---

[1] The author wishes to express his gratitude to Allen Parducci and Norman H. Anderson for comments and suggestions on earlier drafts of this paper. Support was received from National Institutes of Health Grant HD 00923; computing assistance was received from Campus Computing Network, University of California, Los Angeles. Thanks are due to Leslie Freedman and Clairice Veit for their assistance with Exp. I and to Andrew Comrey for providing the successive intervals scaling computer program.

[2] Requests for reprints should be sent to Michael H. Birnbaum, Department of Psychology, University of California, Los Angeles, California 90024.

TABLE 1

Morality Values for the 16 Component Items

| Item no. | Item | Mean judgment | *SD* | Thurstone value | Thurstone dispersion |
|---|---|---|---|---|---|
| 1 | Keeping a dime you find in a telephone booth.[b,d] | 8.79 | .68 | 5.50 | 1.81 |
| 2 | Playing poker on Sunday.[a,c] | 8.78 | .84 | 5.37 | 1.79 |
| 3 | Cheating at solitaire.[c] | 7.89 | 1.58 | 3.66 | 1.54 |
| 4 | Wearing shorts on the street where it is illegal.[d] | 7.81 | 1.43 | 3.38 | 1.50 |
| 5 | Stealing a loaf of bread from a store when you are starving.[d] | 7.75 | 1.73 | 3.52 | 1.52 |
| 6 | Registering in a hotel under a false name.[c] | 7.70 | 1.83 | 3.49 | 1.51 |
| 7 | Stealing towels from a hotel.[b,c] | 7.25 | 1.65 | 2.91 | 1.43 |
| 8 | Failing to pay your bus fare when the conductor overlooks you.[a,c] | 7.04 | 1.74 | 2.78 | 1.41 |
| 9 | Failing to put back in the water lobsters which are shorter than the legal limit.[d] | 6.54 | 2.03 | 2.44 | 1.34 |
| 10 | Getting your own way by playing on other people's sympathies.[b,d] | 5.78 | 2.05 | 2.08 | 1.31 |
| 11 | Contributing money to a cause in which you do not believe in order to escape criticism.[a,d] | 5.33 | 2.27 | 1.87 | 1.28 |
| 12 | Pocketing the tip the previous customer left for the waitress.[c] | 4.85 | 2.04 | 1.72 | 1.25 |
| 13 | Bawling out servants publicly.[b,d] | 4.58 | 1.96 | 1.50 | 1.22 |
| 14 | Habitually borrowing small sums of money from friends and failing to return them.[a,c] | 4.32 | 1.77 | 1.40 | 1.21 |
| 15 | Spreading rumors that an acquaintance is a sexual pervert.[b,c] | 3.27 | .98 | .85 | 1.13 |
| 16 | Poisoning a neighbor's dog whose barking bothers you.[a,d] | 1.95 | 1.46 | .00 | 1.00 |

[a] First item of Exp. I.
[b] Second item of Exp. I.
[c] First item of Exp. II.
[d] Second item of Exp. II.

exceptions to the model have been obtained with psychophysical stimuli (Birnbaum, Parducci, & Gifford, 1971; Parducci, Thaler, & Anderson, 1968). Birnbaum et al. (1971) proposed a range model to account for interactions obtained in these experiments.

The range model asserts that the psychological impression of the stimulus combination is directly related to the *range* of the stimuli as well as the *mean*. For two stimuli, $i$ and $j$, the model can be written:

$$\Psi_{ij} = ws_i + (1 - w)s_j + \omega|s_i - s_j|/(s_i + s_j), \quad [3]$$

where $\Psi_{ij}$ is the impression of the pair, $w$ and $(1 - w)$ are the mathematical weights associated with order of presentation, $s_i$ and $s_j$ are the scale values of the stimuli, and $\omega$ is an empirical constant which represents the magnitude of the range effect. The scale values are assumed to all have the same sign.

The present experiments provide a test of the applicability of the general additive model to morality judgments and permit a comparison with the range model.

## METHOD

In both experiments, *S*s were instructed to read each pair of behavior items and judge "how wrong that *pair* of actions would be in terms of your own personal set of values."

*Behavior items.*—Brief characterizations of 16 immoral behaviors were selected from those used by Parducci (1968). These single actions were judged by 81 undergraduates who made their ratings on a 9-point scale. The category labels were the same as in Exp. II below. The mean category judgments are presented as the scale values in Table 1, along with the corresponding Thurstone Case 6 values (Bock & Jones, 1968).

*Subjects.*—The *S*s were 219 University of California, Los Angeles, undergraduates; 81 *S*s judged the single items for the preliminary scaling. There were 100 *S*s in Exp. I and 38 different *S*s in Exp. II. The *S*s were run in groups of from 2 to 10 *S*s each.

*Experiment I.*—The stimuli consisted of 25 pairs of items produced from a 5 × 5 factorial design in which the first and second items could each take five values. Table 1 lists these items. The 25 pairs were printed in booklets in a single random order, and *S*s were instructed to read the entire list before making their judgments.

The ratings were recorded in numerical form, using a 9-point scale anchored by five category labels: 1—"neither good nor bad," 3—"questionable, but not particularly bad or wrong," 5—"undesirable, a good person would not do this," 7—"wrong, highly questionable," 9—"seriously wrong." To facilitate comparison with related work, judgments obtained in both experiments were subtracted from 10 to reverse the scale.

*Experiment II.*—The two experiments differed in the labeling of the response scale and also in the particular stimulus pairs. The labels were adjusted for Exp. II to permit more categories at the "seriously wrong" end of the scale. The new labels were: 1—"not particularly bad or wrong," 3—"undesirable, a good person would not do this," 5—"wrong, highly questionable," 7—"seriously wrong," 9—"extremely evil." The stimuli of Exp. II consisted of 64 pairs of items, produced from an 8 × 8 factorial design using all 16 items listed in Table 1. Each *S* received a booklet in which the pairs were printed in one of four random orders. The *S*s judged each pair twice; the order of the pairs and the order of items within each pair was reversed for the second replicate. The order of items within each pair on the first replicate was counterbalanced on half of the forms.

## RESULTS AND DISCUSSION

Figure 1 plots the mean rating obtained in both experiments against the best-fit values of the averaging model. The model is fitted separately for each experiment by a least-squares solution for the weights and additive constant $(c + w_o s_o)$ of Equation 1, using the prior scale values of Table 1. The multiple correlations for Exp. I and II were .938 and .952, respectively. These high coefficients indicate that the additive (or averaging) model provides a satisfactory basis for practical prediction, and they also reflect high reliability of the scaling. However, any model predicting that response varies directly with each of the component values would yield high correlations (Anderson, 1968b; 1971). Therefore, it is necessary to examine the deviations from the model.

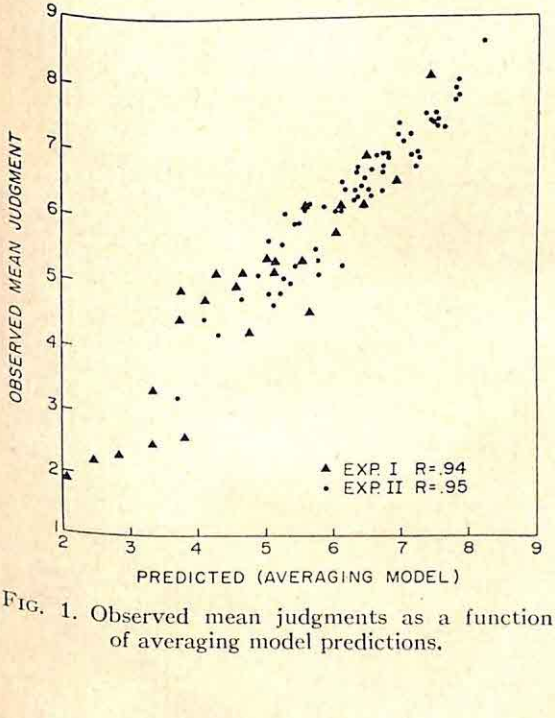


FIG. 1. Observed mean judgments as a function of averaging model predictions.

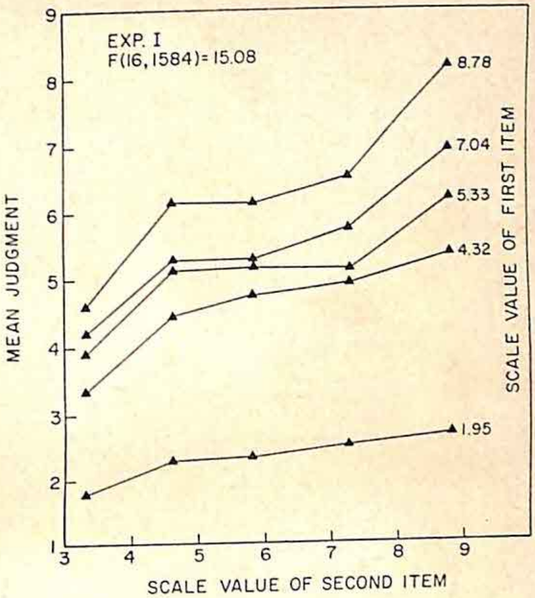


FIG. 2. Mean judgment of each pair of behaviors of Exp. I as a function of the scale values of the first and second items.

Figure 2 shows the mean judgments of the 25 pairs of Exp. I. Each curve represents the mean judgments of five pairs with the same first item. The slope of each curve represents the effects of the second item; the distances between the curves represent the effects of the first item. According to additive (or averaging) models, the curves should be parallel. Instead, they diverge to the right. This divergence is characteristic of the individual data for 85 of the 100 *S*s. The analysis of variance test of the interaction is highly significant, $F\ (16, 1584) = 15.08, p < .001$. This interaction is inconsistent with the

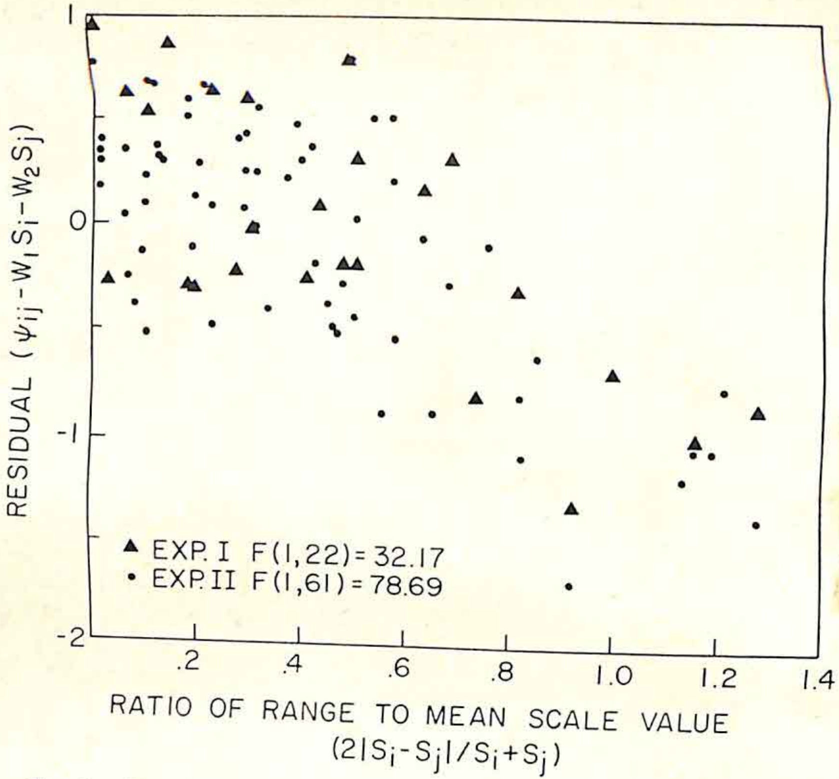


FIG. 3. Residuals from the best fit of the averaging model (following transformation) as a function of the ratio of the within-set range to the mean scale value of the pair.

general additive model (Equation 1) and therefore provides evidence against all of the special cases of this model (Rosenberg, 1968).

One possibility is that these interactions (deviations from additivity) reflect compression of the response scale at the "seriously wrong" end, i.e., nonlinearity in the function *J* of Equation 2. As a check on the possibility of this "end" effect, all pairs containing either of the two items with the lowest scale values were removed from the analysis, which was thus reduced to a 4 × 4 design. The interaction remained highly significant, $F(9, 891) = 8.35$, $p < .001$.

The data of Exp. II provided a second check on the response scale. Although the change in scale labels for Exp. II reduced the use of the lower categories (as shown in Fig. 1), the interaction had the same form as that of Exp. I and was highly significant, $F(49, 1813) = 4.21$, $p < .001$.

## *Monotonic Transformation*

As a final check on the response scale, the responses were transformed to impression values ($\Psi$ in Equation 2) by the procedure described in Birnbaum et al. (1971). This procedure assumes the validity of the prior scale values and the validity of the averaging model for pairs of minimal within-set range. The transformation, which relates the judgments of pairs to impression values on the prior scale, is then estimated as a polynomial by a least-squares criterion. Although this procedure permits radical rescaling, the best-fit transformation was fairly linear so that the interactions retained the same form and remained highly significant. The finding that the ratings of pairs of minimal within-set range are linearly related to the ratings of the single items suggests that the failure of the additive and averaging models is not due to improper scaling of the responses.

### *Range Model*

Figure 3 plots the residuals from the additive model of Equation 1 (following transformation) as a function of the ratio of the range to the mean scale value for each pair. Since these residuals are simply "obtained minus predicted," they should be zero in theory and at least show no trend. Instead, the relationship follows the linear trend predicted by the range model (Equation 3). In addition, the transformed data from both experiments yield approximately the same estimate of $\omega$ for Exp. I and II ($-2.61$ and $-2.32$); i.e., both sets of points fall approximately on the same line. The partial correlations with the range term, $-.77$ and $-.75$ for the two sets of data are highly significant, $F\,(1, 22) = 32.17$ and $F\,(1, 61) = 78.69$. The sum of squared errors for the additive or constant-weight averaging models is more than twice the sum of squared errors for the range model.

### *Theoretical Interpretations*

*Overlap of values.*—A possible theoretical basis for the range model assumes that each component stimulus produces a distribution of values on an underlying psychological continuum. This hypothetical distribution represents all the possible values of the component stimulus. The present hypothesis assumes that *S* bases his rating of a single component on the mean of its distribution of possible values; when instructed to judge a pair, he estimates the mean of those values common to both distributions, i.e., their overlap. If the dispersion of each component's distribution of possible values varies directly with its mean, the value of the combination will be related to the range (i.e., difference between the means of the components) as well as to the average of the two mean values. As shown in Fig. 4, the rating of the combination will be shifted toward the component with narrower dispersion; the relative magnitude of the shift increases with the difference in dispersion. For the present data, this hypothesis requires that the immoral items have narrower dispersions than the milder items.

Application of successive interval scaling (Torgerson, 1958, pp. 205–210) to the ratings of the 16 single items revealed that scale value and dispersion were indeed correlated ($r = .82$). This correlation suggested the use of Thurstone Case 6 values in Fig. 4 to illustrate the hypothesis. Since the function relating Thurstone values to mean judgments is negatively accelerated, the effect of the difference in dispersion is inversely proportional to the mean scale value.

Range effects have been reported in previous studies of information integration. Willis (1960) found that judgments of the attractiveness of groups of facial photographs varied inversely with the "heterogeneity" of the component photograph scale values and directly with their mean. Heterogeneity was defined as the average absolute deviation of the components around their mean and is therefore analogous to the range term of Equation 3. Weiss (1963) found similar interactions for judgments of opinion triplets. Judgments of intelligence show interactions due to "cue inconsistency," a concept which is also analogous to that of range (Slovic & Lichtenstein, 1971).

The range model has also been applied to judgments of the likableness of hypothetical persons described by pairs of adjectives, including adjectives from both ends of the likableness continuum. The range term accounted for 80% of the variance left unexplained by the additive models (Birnbaum, 1970, 1971). Work is currently under way to investigate the relative effects of moral and immoral items on integrated impressions of morality.

*Other interpretations.*—Another interpretation of the interactions postulates shifts in the components themselves. When separate ratings of the components presented in the set (Anderson, 1968b, 1971; Birnbaum et al., 1971) shift systematically, it is usually toward the values of the other components of the same set. However, if the magnitude of shifts were simply a linear function of the values of the com-

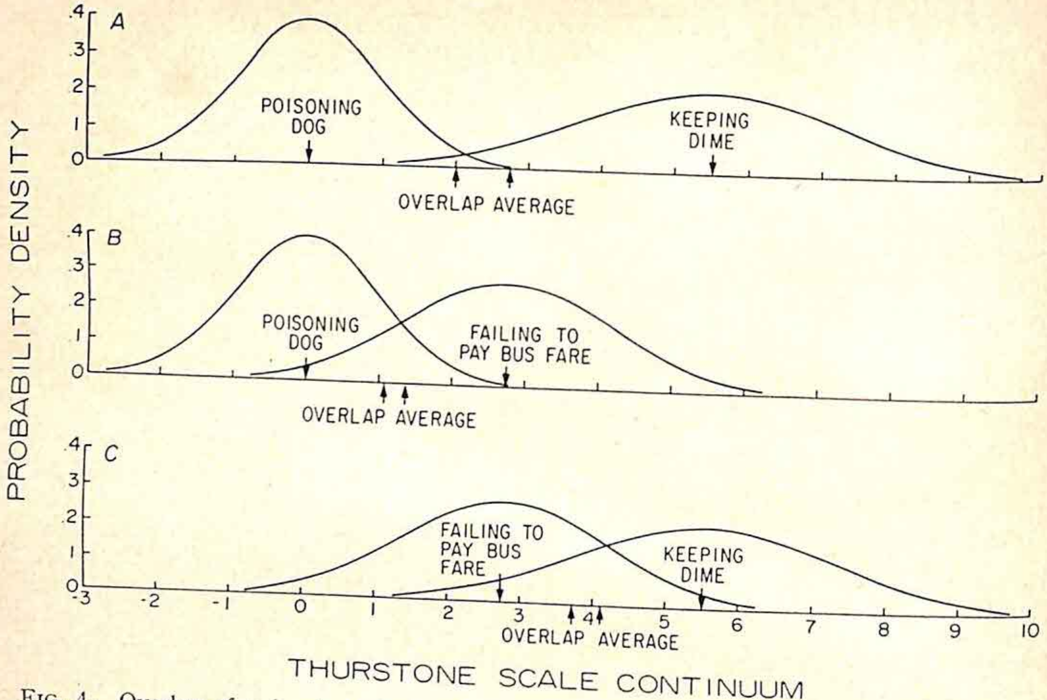


FIG. 4. Overlap of value hypothesis illustrated. (Arrows labeled "average" represent the averaging model predictions; each arrow labeled "overlap" represents the mean of the overlap of two distributions.)

ponents, this contextual effect would not produce interactions (Anderson, 1968b, 1971).

Instead of the items changing value in context, it is possible that the weight or importance of each component depends upon its value (Anderson, 1971) and perhaps also on the values of the other components of the same set. Osgood and Tannenbaum (1955) have suggested a model in which the weight of an item is proportional to its extremity. Differential weighting has also been considered by Manis, Gleason, and Dawes (1966) and by Feldman (1968). There are many possible explanations for differential weighting models. For the present data, the more serious misbehaviors might receive greater relative weight in determining the overall impression because they require action such as punishment of the wrongdoer. Similarly, if serious misbehaviors occur less often, these items would provide more information and consequently receive greater weight. Weighting could also depend upon the reliability of the information, with weight inversely proportional to dispersion (as in Fig. 4).[3]

Differential weighting models might be consistent with the present interactions which contradict the constant-weight models. These more complicated averaging models require the estimation of two parameters for each stimulus, weight and scale value, whereas the constant-weight models (with appropriate experimental designs) require the explicit estimation of only the scale values. The number of parameters can be reduced by assuming that weight is a linear function of scale value. This simplified variable-weight model does as well as the range model for the data of Exp. I, but poorer for Exp. II. After fitting this variable-weight model to the data of Exp. II, the range term accounts for a significant portion of the unexplained variance, $F\ (1, 60) = 22.6, p < .001$.

---

[3] The distributional interpretation shown in Fig. 4 can provide the basis for the derivation of an averaging model with differential weights. Let $s_i$ and $\sigma_i$ be the mean and standard deviation for Stimulus $i$. If it is assumed that the combined impression, $\Psi$, of $k$ stimuli is the value that maximizes the product of the normal probability densities (as in Fig. 4), then it can be shown that $\Psi = \sum_{i=1}^{k} (s_i/\sigma_i) / \sum_{i=1}^{k} (1/\sigma_i)$, which is an averaging model with weights inversely proportional to dispersion. For two stimuli, this criterion also implies that $P(M > \Psi | \text{Stimulus } 1) = P(M < \Psi | \text{Stimulus } 2)$, where $M$ is morality.

The distinction between differential weighting models and the constant-weight additive or averaging models cannot be emphasized too strongly. The additive models (Rosenberg, 1968) require many fewer parameters but cannot account for interactions. The differential weighting models can account for interactions (Anderson, 1971) but suggest a different process of information integration and raise the theoretical problem of explaining the nature and cause of the weighting.

On the other hand, it would be difficult at present to make a sharp distinction between the range model and the differential weighting models on the basis of existing data, since both models can predict simple convergent or divergent interactions. For example, recent experiments by Oden and Anderson (1971) suggest that judgments of the "badness" of groups of criminals are inconsistent with additive models of information integration. A differential weighting model was applied to the data, although the deviations from additivity were in the direction that would be predicted by the range model.

## CONCLUSIONS

The present data suggests that *S*s do not combine information about the morality of behaviors by simply summing or averaging the values associated with the components. The interactions obtained in both experiments are consistent with other evidence (e.g., Anderson, 1965, 1968a; Feldman, 1968; Lampel & Anderson, 1968; Oden and Anderson, 1971) suggesting that integrated impressions are shifted toward the evaluatively lower information contained in the set.

The present experiments do not preclude the possibility that the interactions are due to contextual changes in value or weight. Differential weighting of the stimuli may account for the shifts but the range model appears to provide a better fit to the data.

The range model can be interpreted as reflecting an underlying psychological representation of the stimulus items in which the lower valued stimuli are located more precisely than the neutral ones, so that the stimulus combination falls closer to the lower items.

## REFERENCES

ANDERSON, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, 1965, **70**, 394–400.

ANDERSON, N. H. Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology*, 1968, **10**, 352–362. (a)

ANDERSON, N. H. A simple model for information integration. In R. P. Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. (b)

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

ANDERSON, N. H. Information integration and attitude change. *Psychological Review*, 1971, **78**, 171–206.

BIRNBAUM, M. H. Impression formation: Averaging versus range model. Paper presented at the Mathematical Psychology Meeting, Miami, September 1970.

BIRNBAUM, M. H. Impression formation: Difference judgments as a basis for response rescaling. Paper presented at the meeting of The Western Psychological Association, San Francisco, April 1971.

BIRNBAUM, M. H., PARDUCCI, A., & GIFFORD, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, **88**, 158–170.

BOCK, R. D., & JONES, L. V. *The measurement and prediction of judgment and choice.* San Francisco: Holden-Day, 1968.

FELDMAN, S. What do you think of a *cruel, wise* man? The integrative response to a stimulus manifold. In R. P. Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

LAMPEL, A. K., & ANDERSON, N. H. Combining visual and verbal information in an impression formation task. *Journal of Personality and Social Psychology*, 1968, **9**, 1–6.

MANIS, M., GLEASON, T. C., & DAWES, R. M. The evaluation of complex social stimuli. *Journal of Personality and Social Psychology*, 1966, **3**, 404–419.

ODEN, G. C., & ANDERSON, N. H. Differential weighting in integration theory. *Journal of Experimental Psychology*, 1971, **89**, 152–161.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, **62**, 42–55.

PARDUCCI, A. The relativism of absolute judgments. *Scientific American*, 1968, **216**, 84–90.

PARDUCCI, A., THALER, H., & ANDERSON, N. H. Stimulus averaging and the context for judgment. *Perception & Psychophysics*, 1968, **3**, 145–160.

ROSENBERG, S. Mathematical models of social behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. Vol. 1. (2nd ed.) Reading, Mass: Addison-Wesley, 1968.

SLOVIC, P., & LICHTENSTEIN, S. C. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, **6**, 649–744.

TORGERSON, W. *Theory and methods of scaling*. New York: Wiley, 1958.

WEISS, W. Scale judgments of triplets of opinion statements. *Journal of Abnormal and Social Psychology*, 1963, **66**, 471–479.

WILLIS, R. H. Stimulus pooling and social perception. *Journal of Abnormal and Social Psychology*, 1960, **60**, 365–373.

(Received June 18, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 43–55

# HAPTIC JUDGMENTS OF CURVATURE BY BLIND AND SIGHTED HUMANS[1]

PHILIP W. DAVIDSON[2]

*George Washington University*

Three experiments investigated the relationship between active handling and veridical haptic curvature perception. Experiment I showed that blind *S*s made more objective judgments of curves than did sighted *S*s. Videotape recordings of exploratory scanning during judgments revealed that the blind used a scanning technique allowing more global apprehension of the stimulus than sighted *S*s' methods, and that the two groups focused on different stimulus parts. In Exp. II, sighted *S*s' judgments became more objective when they were restricted to using the scanning technique characteristic of the blind. Experiment III suggested that sighted *S*s' errors in Exp. I probably stemmed from using movements that obscured stimulus curvature. It was concluded that the blind's scanning technique (*a*) focused attention to an informative stimulus feature, and (*b*) provided a way to pattern the ends-to-middle relationship that specifies curvature. Implications for a theory of the role of exploratory activity in perception were discussed.

Judgments about the characteristics of an object typically call for exploratory activity—eye or hand movements from one feature to another. When the judgment is visual, the object is explored by shifting the direction of gaze, whereas haptic judgments involve feeling different stimulus parts with the fingers and hands.

Exploratory activity has been related to perceptual skill in several recent theories of perception. Gibson (1962), for example, suggested that haptic exploratory movements "isolate and enhance the component of stimulation that specifies the shape and other characteristics of the object being touched [p. 478]." Similarly, Zinchenko and Lomov (1960) summarize the Russian view that eye and hand movements serve to organize the parts of a form into a whole. Finally, changes in exploratory activity, linked with improved perceptual skill during ontogenetic development (cf. Abravanel, 1968; Nodine & Lang, 1971; Zinchenko & Ruzskaia, 1960a, 1960b), and with experience using touch in blind *S*s (cf. Hatwell, 1959, 1960), suggest that the accuracy of information gathered about an object may be greatly influenced by the way the object is explored.

Despite the important role assigned to exploratory activity by theoretical views, the nature of its contribution to perception is relatively unknown. Some exploratory methods lead to greater accuracy of judgment than others. When the judgment is of form, for example, eye and finger movements that focus attention to the object's contour seem related to accurate identifications (Abravanel, 1968; Zinchenko & Ruzskaia, 1960a, 1960b). The way the contour is explored too is important: With touch, tracing around a contour with one finger is less successful with respect to accurate judgments than to feel many points on the contour at once with all of the fingers (Abravanel, 1968; Hatwell, 1959, 1960; Zinchenko & Ruzskaia, 1960a, 1960b). In addition, certain techniques seem best suited for exploring particular kinds of stimuli, or stimulus attributes. For example, Abravanel found that clutching the ends of a cylinder with the hands was a satisfactory exploratory method for accurately perceiving diameter, but inappropriate for apprehending the length of a bar.

---

[1] This research is based on the author's doctoral dissertation, submitted to George Washington University. The study was supported by Public Health Service Predoctoral Fellowship MH-42881, and a National Science Foundation Summer Research Fellowship, both to the author, and by National Institute of Child Health and Human Development Grants 1-T01-HD00203 to Richard D. Walk, and 1-K03-HD36971 to Jacqueline J. Goodnow. The author wishes to thank R. D. Walk, J. J. Goodnow, and E. Abravanel for their help in all phases of the project, and M. Smith and C. Walters for scoring the videotaped data.

[2] Requests for reprints should be sent to Philip W. Davidson, Department of Psychology, Washington College, Chestertown, Maryland 21620.

These data suggest that the contribution of exploratory activity to perception probably lies in some interaction between the nature of the stimulus and the nature of the scanning. However, with few exceptions, little in the way of hypotheses or data has been reported to describe the interaction.

The present experiments, designed to pin down the interaction between stimulus properties and scanning, deal with the relationship between one haptic form judgment and haptic exploratory activity. Active touch was selected for study because scanning methods can be readily observed and changes in method easily detected. In addition, it is often easy with touch to relate particular scanning methods to particular aspects of the stimulus.

The task calls for judging the curvature of an edge by hand. This is one of the few tasks that human blind *S*s perform more objectively than the sighted (Hunter, 1954; Pick & Pick, 1966). There is a hint that the blind *S*s' objectivity may stem from having acquired new ways of exploring, illustrating a general form of the interaction between the type of scanning and the stimulus.

The design consists of two main steps. First, the methods blind and sighted *S*s used to feel curved stimuli are compared, suggesting differences in scanning that might be critical to accuracy (Exp. I). Then, the importance of these aspects of scanning to objective judgments is tested by introducing constraints, either by restricting scanning methods (Exp. II), or by altering characteristics of the stimulus (Exp. III).

## EXPERIMENT I

Hunter (1954) had blind and sighted *S*s explore by touch alone a flexible ruler standing on its long edge, sometimes bent to describe different concave and convex symmetrical arcs in the transverse plane, and sometimes left straight. The task was to judge whether the ruler felt straight, convex, or concave. Hunter found that the stimulus felt "straight" to both blind and sighted *S*s when it was actually concave (ends bent toward *S*), but that subjective straightness for the blind was a significantly less acute concave curve than for sighted *S*s.

The Hunter (1954) experiment leaves us very uncertain regarding the nature of haptic curvature perception by either blind or sighted *S*s. First of all, Hunter's data reflect only a determination of psychophysical "straightness," neglecting to report *S*s' performance along the rest of the curvature continuum. Moreover, the reason for the more objective judgments by blind *S*s is unclear. Other studies of haptic curvature perception have indicated that a curve feels different depending upon how it is explored (Goodnow, Baum, & Davidson, 1971; Rubin, 1936). It may be that blind *S*s' superior performance in Hunter's (1954) experiment stems from having learned more effective methods for scanning the stimulus.

Experiment I categorized the scanning techniques *S*s use to feel curves like those used by Hunter (1954) and evaluated differences in scanning style between the blind and the sighted. The study also extended Hunter's "straightness" comparison, asking if blind and sighted *S*s differ in perception of individual convex and concave curves, and a straight edge.

### *Method*

*Subjects.*—Thirty-two male and female *S*s (nearly all right-handed) participated in the experiment. Half were normally sighted undergraduates. The other half were totally blind from birth, with blindness stemming exclusively from retrolental fibroplasia. The blind *S*s were selected from Grades 10–12 of the Maryland School for the Blind,[3] and the average age was 18 yr., 0 mo.

*Stimuli.*—Five stimuli replicated most of the curvature continuum used by Hunter (1954). Four of the stimuli were symmetrical curves as shown in Fig. 1, and the fifth was a straight stimulus. Each stimulus consisted of a 200-mm. long, 19-mm. high, 1.5-mm. thick plastic strip, cemented to the curved edge (or straight edge in the case of the straight stimulus) of a heavy plastic base. The arc heights of the four curved stimuli (the perpendicular distance from the midpoint of the arc's chord to the midpoint of the arc) were 2, 4, 6, and 8 mm.

---

[3] The author gratefully acknowledges the cooperation of the administration, faculty, and staff of the Maryland School for the Blind.

Figure 1 also shows how stimuli were positioned. A curve could be placed in the holder oriented either as a concave arc (ends bent toward *S* as in Part A of the figure) or as a convex arc (ends bent away from *S* as in Part B).

*Videotape equipment and other materials.*—The exploratory activity of each *S* was recorded on videotape with a Sony videotape system. The videocamera was tripod-mounted and equipped with a 75-mm. lens to enlarge the image.

All *S*s wore a blindfold during testing. An ashtray covered with foam rubber served as an elbow rest.

*Task.*—The task was to categorize stimuli as convex, concave, or straight, following a "Method of Single Stimuli" procedure (cf. Guilford, 1954). The *S* felt each of the four curved stimuli eight times—four times as a concave curve and four times as a convex curve—and he felt the straight stimulus four times. The stimulus order was randomized for the 36 judgments.

The stimuli were positioned in front of *S* so that the edges were at fingertip distance. The *S*s explored with their preferred hand, keeping their elbows in the elbow rest.

After the task was explained, *S* explored a 16-mm. warm-up curve, oriented once as concave and once as convex, until he correctly categorized each orientation.

*Videotape recording.*—The *S*'s exploratory activity during all judgments was recorded on videotape. The camera, positioned about 2.5 m. to the left front of *S*, looked downward onto the stimulus. The oblique angle gave a clear view of wrist and finger movements and of the thumb of a right-handed *S*.

*Videotape scoring.*—Two scorers, trained by *E*, analyzed the videotapes. The scorers observed each taped judgment noting the appearance of any of a number of predefined exploratory scanning techniques. Each scorer viewed the tapes independently and was unaware of *S*s' identities.

The videotapes were scored twice. The first scoring was for five aspects or "elements" of scanning, defined as follows: (*a*) Part of curve touched (back edge, i.e., the edge facing *S*, top edge, or front edge). (*b*) Number of fingers used (one, two, three, or four, excluding the thumb). (*c*) Position of fingers (scored as stiff or bent at the knuckles, with the stiff category subdivided into two possibilities: fingers held parallel to the table or at an acute angle). (*d*) Thumb placement on back edge of stimulus (scored as used in a pincer with fingers touching other parts of the curve, used independently of other fingers, or not used). (*e*) Style of movement (scored as smooth and sweeping, uneven and sporadic, or no movement).

The second scoring of the tapes was to see if the scorers could distinguish among some general methods of scanning called "classes" that resulted from different combinations of elements. Five scanning classes—GRIP, PINCH, FINGERTIP SWEEP, SPAN, and TRACE—were defined by *E* after a study of the videotapes and are depicted in Fig. 2. The grip involved curling three or four fingers over the front edge of the stimulus, pulling back slightly against it, while shifting back and forth across the curve (Fig. 2A). The pinch (Fig. 2B) consisted of holding the stimulus edge in a pincer between the thumb and one or more other fingers, sliding smoothly across the curve. The fingertip sweep took two forms. The top sweep (Fig. 2C) consisted of placing one or more fingers, held straight out and together, on the top edge of the curve and sweeping back and forth. A variant of this method was the front sweep, where *S* scanned the front edge as with the top sweep. A majority of *S*s performed the top sweep with two fingers. The span (Fig. 2D) involved spanning the curve with the whole hand, fingers spread apart and outstretched. The trace (Fig. 2E) consisted of holding one or two fingers perpendicular to the table, scanning across the front edge to the end of the curve, then scanning in the opposite direction across the back edge.

*Training the scorers.*—Three 1-hr. training sessions preceded the scoring. In each session, *E* presented various scanning techniques on the video monitor by photographing his own hand, while the scorers practiced noting the element categories and classes.

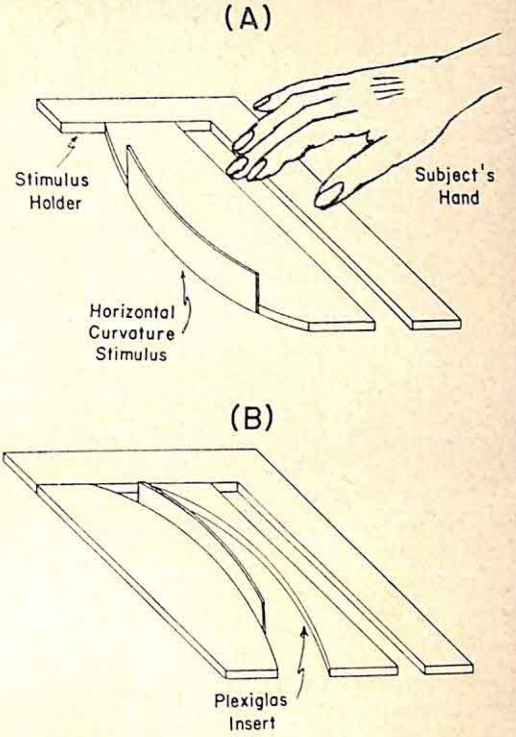


FIG. 1. Stimulus curves for Exp. I. (The upper figure, A, shows a stimulus oriented as a concave curve and the lower figure, B, shows a convex curve.)

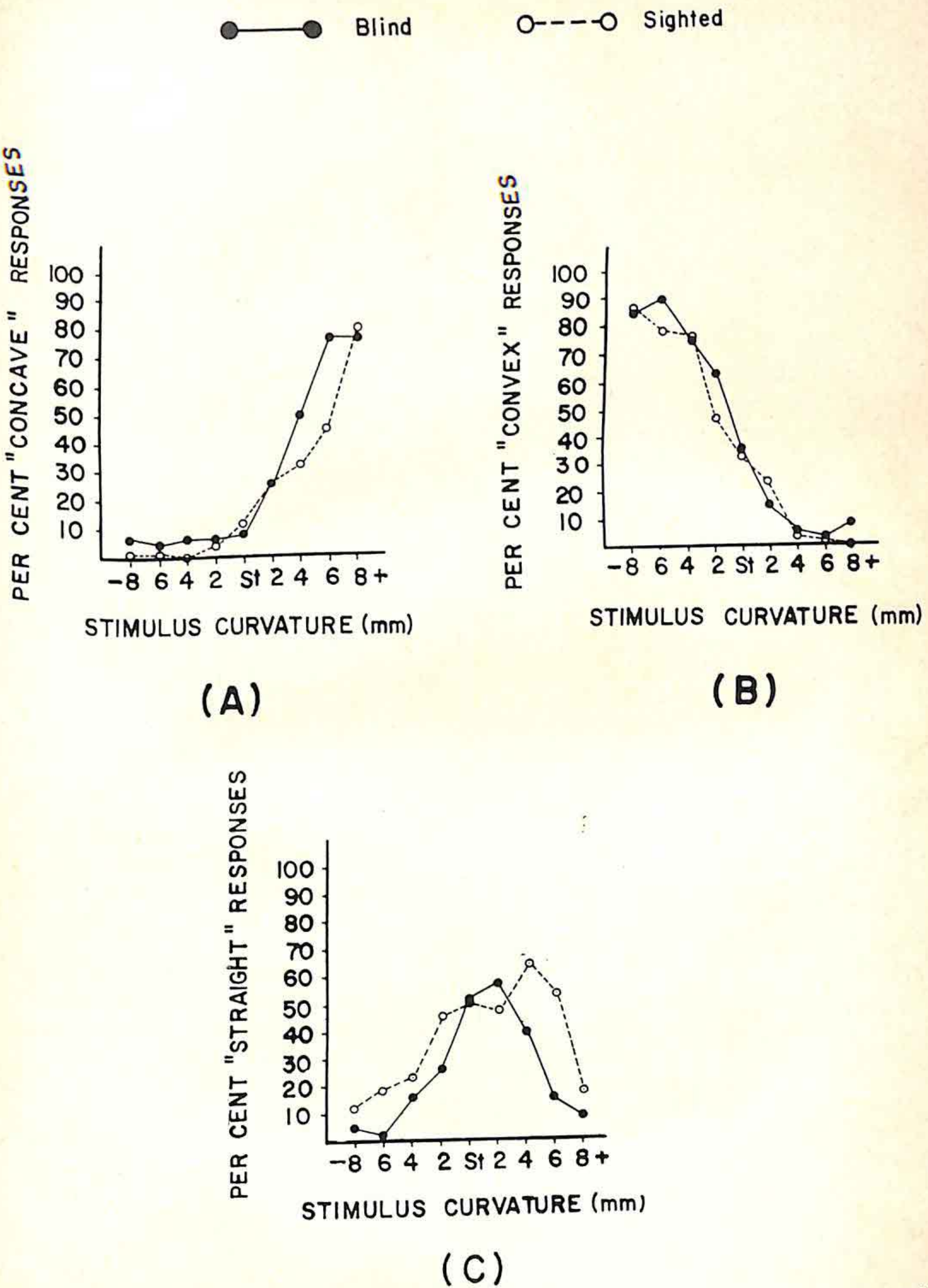


FIG. 3. How *S*s categorized stimuli in Exp. I. (The figure shows the frequency of "concave," A, "convex," B, and "straight," C, responses by each group of *S*s to each curvature value. Concave stimulus values are represented on the abscissa as "+," convex values as "−.")

and 3B, taken together, suggest that the frequencies of "concave" and "convex" responses to the straight stimulus were low, and similar for both groups (around 25%).

Psychophysical scales of curvature for

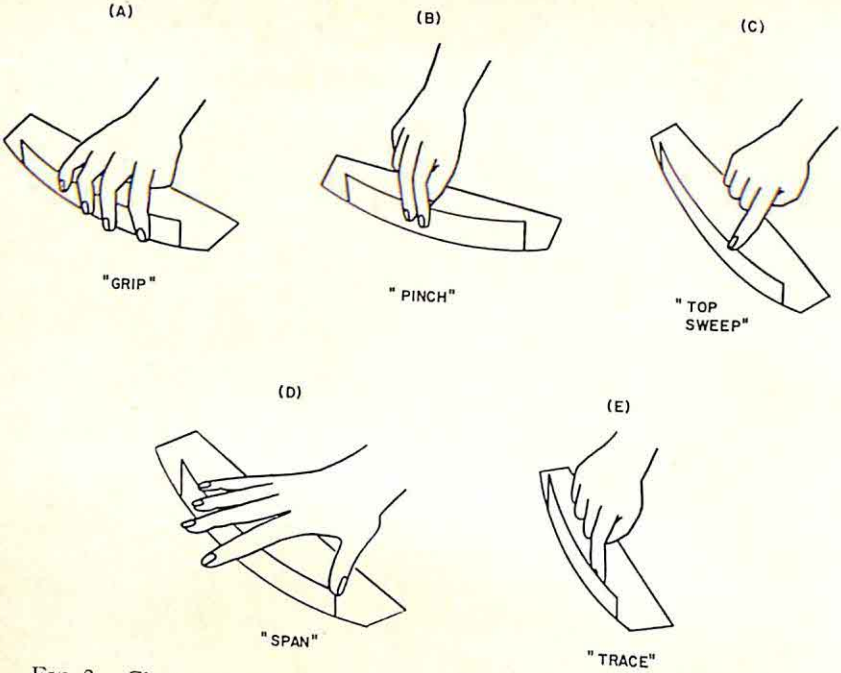


Fig. 2. Classes of scanning. (The figure shows the active handling strategies that *S*s used to explore curves in Exp. I. The grip, A, pinch, B, and top sweep, C, were the most frequently used techniques.)

The *E* also emphasized critical differences among the scanning methods. Brief segments of the data tapes were randomly selected from many *S*s' judgment series and shown during the final training session.

*Results*

Blind *S*s gave significantly more objective judgments than sighted *S*s. Both groups' incorrect categorizations occurred most often with concave stimuli, but blind *S*s showed more correct categorizations of those curves than sighted *S*s. Both groups correctly categorized convex curves, and incorrect responses to the straight stimulus for both groups were distributed equally between "concave" and "convex." In addition, Exp. I indicated (*a*) significant differences between blind and sighted *S*s in frequency of occurrence of certain scanning aspects and (*b*) the predominant use by the blind of a particular scanning technique—the grip—infrequently observed in judgments of sighted *S*s.

*Objectivity of curvature judgments.*—The mean number of incorrect categorizations provides an index of overall group performance. Blind *S*s made an average of 11.44 ($SD = 5.45$) incorrect categorizations, significantly less than 14.94 ($SD = 4.82$), the average for the sighted *S*s, $t\ (30) = 1.92$, $p < .05$. In addition, 57% of the incorrect categorizations for each group occurred with concave curves.

Figure 3 gives the main data from Exp. I, showing how *S*s responded to each value along the curvature continuum studied. Figure 3A shows the frequency of "concave" responses by each group of *S*s to each curvature value, 3B shows the frequency of "convex" responses, and 3C, the frequency of "straight" responses. Figure 3A shows that blind *S*s gave more "concave" responses than sighted *S*s to the 4-mm. and 6-mm. concave curves, a difference mirrored in Fig. 3C, where sighted *S*s exceeded the blind in frequency of "straight" responses to these two stimuli. In contrast, Fig. 3B shows an extremely high frequency of "convex" responses to convex stimuli by both blind and sighted *S*s and indicates that neither group categorized either the 4-mm. or 6-mm. concave curves as "convex" more than 2% or 3% of the time. Finally, Fig. 3A

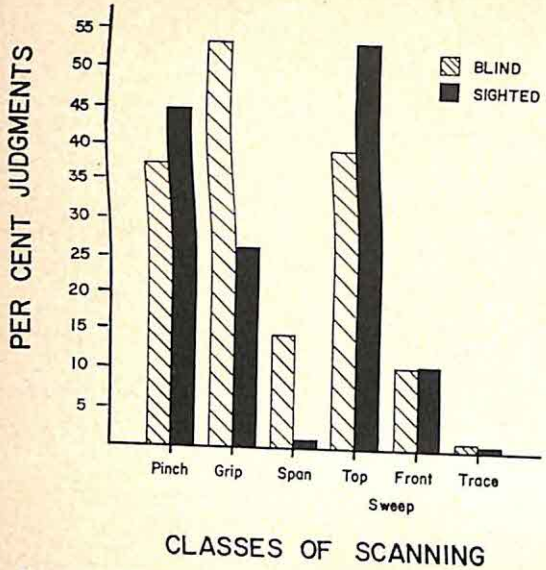


FIG. 4. The percentage of judgments in which blind and sighted *S*s used the Scanning Classes. (Each *S* made 36 judgments. Mann-Whitney *U* tests indicated that blind *S*s significantly exceeded the sighted in frequency of use of the grip, $p < .02$, and the span, $p < .10$.)

blind and sighted *S*s were derived following a least-squares solution for the Method of Single Stimuli (as described by Guilford, 1954, p. 128). This procedure, summarized in Table 1, uses group percent "concave" (Fig. 3A) and percent "convex" (Fig. 3B) response distributions to yield absolute thresholds (AL) for "concavity" and "convexity" and to estimate a point of subjective equality (PSE) for phenomenal straightness. The table shows that the curvature scales for both blind and sighted *S*s are skewed in the direction of concavity and that the skew is greater for sighted *S*s than for the blind. The PSE for blind *S*s is slightly less than a 1-mm. concave curve whereas the sighted *S*s' PSE is nearly a 2-mm. concave curve. Also, the ALs for concavity for both groups occur well into the concave stimulus range, and the AL for sighted *S*s (5.09 mm. concave) is nearly 2 mm. further along the concave range than the AL for blind *S*s (3.27 mm. concave). Finally, the variability associated with the PSE (as measured by the size of the interval of uncertainty, IU) is substantially greater for sighted *S*s than for the blind. The findings regarding phenomenal straightness seem to correspond well to the comparison of distributions of "straight" responses for blind and sighted *S*s shown in Fig. 3C.

*Analysis of exploratory scanning.*—The scorers' task was to classify scanning techniques in each of *S*'s 36 judgments. After checking for interscorer agreement (the scorers agreed approximately 75% of the time[4]), we calculated from these data the number of *S*'s judgments in which the scorers said they observed a particular scanning technique. (When the scorers disagreed on the number of judgments in which a method occurred, their scores were averaged). For each scanning technique, these data were expressed as a percentage frequency of occurrence of the technique in *S*'s judgments, and averaged for each group of *S*s. The resulting distributions for classes of scanning in blind and sighted *S*s' judgments are compared in Fig. 4. The element categories are compared in Fig. 5. The difference between these distributions for blind and sighted *S*s was tested with separate Mann-Whitney *U* tests (as described in Siegel, 1956) for each scanning class and element category.

Figure 4 shows that the grip, pinch, and

TABLE 1

PSYCHOPHYSICAL SCALES OF CURVATURE FOR BLIND AND SIGHTED *S*s

| *S*s | Curvature thresholds (mm.) | | | | IU (mm.) | PSE (mm.) (Phenomenal straightness) |
|---|---|---|---|---|---|---|
| | Convex (−) | | Concave (+) | | | |
| | AL | *SD* | AL | *SD* | | |
| Blind | −1.27 | 5.50 | +3.27 | 6.09 | 4.54 | + .88 |
| Sighted | −1.65 | 6.03 | +5.09 | 5.73 | 6.74 | +1.80 |

Note.—PSE = Point of subjective equality. IU = interval of uncertainty. AL = absolute threshold.

---

[4] Interscorer agreement is difficult to evaluate statistically when the data generated by the scorers is nominal, and when there are multiple, related categories. The most common procedure is to report the percent frequency of scorer agreement (cf. Abravanel, 1968; Scott & Wertheimer, 1962). The overall agreement (for elements and classes) was 73.88%, with the lowest scores associated with the least used categories. The individual agreement scores for each element category and class are given in Davidson, 1970.

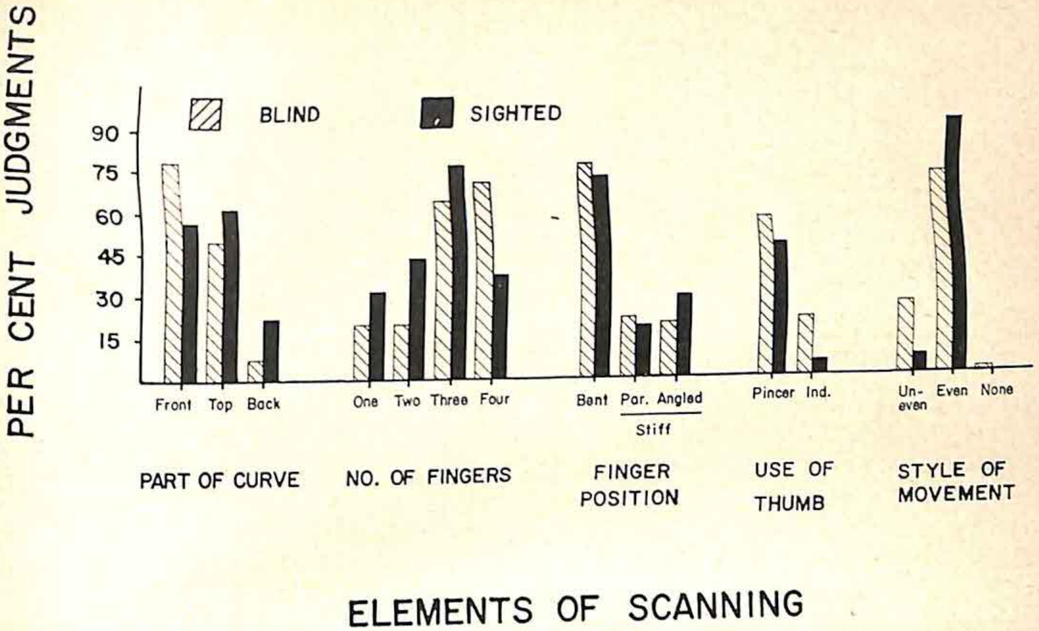


FIG. 5. The percentage of judgments in which blind and sighted *S*s showed each element category. (Each *S* made 36 judgments. Mann-Whitney *U* tests indicated that blind *S*s significantly exceeded the sighted in frequency of use of front-of-curve, $p < .10$, four fingers, $p < .02$, and independent use of the thumb, $p < .05$, while the sighted used two fingers more frequently than the blind, $p < .02$.)

top sweep were the classes of scanning that appeared in the most judgments of both groups of *S*s. One striking finding was that blind *S*s gripped the stimulus in about 60% of their judgments, more often than they used any other class, but the sighted gripped in only about 25% of their judgments, significantly less often than blind *S*s, $U$ (16, 16) = 62.0, $p < .02$. Both seemed to use the pinch and top sweep equally often. Although blind *S*s spanned stimuli more often than the sighted, $U$ (16, 16) = 83.0, $p < .10$, this class appeared in only about 8% of blind *S*s' judgments.

Figure 5 shows that blind *S*s used two fingers significantly less, $U$ (16, 16) = 61.0, $p < .02$, but four fingers significantly more often, $U$ (16, 16) = 54.0, $p < .02$, than the sighted *S*s. Also, blind *S*s explored the front edge of the stimulus more often than the sighted *S*s, $U$ (16, 16) = 80.5, $p < .10$. These differences probably relate to the blind *S*s' more frequent use of the grip, involving exploration of the front edge of the stimulus with three or four fingers.

A comparison of the mean number of scanning classes used per judgment by blind *S*s ($M$ = 1.66) and sighted *S*s ($M$ = 1.55) indicated no significant difference, $t$ (30) < 1.0.

### *Discussion*

Experiment I revealed that categorization errors for both blind and sighted *S*s occur almost exclusively with concave curves and seem to result from a confusion of stimulus concavity with phenomenal straightness. This finding extends previous conclusions (cf. Hunter, 1954, Rubin, 1936) that haptic phenomenal straightness is probably a concave curve. Indeed, we found that the psychophysical scale of haptic curvature seems to be skewed in the direction of concavity, at least when the curvature is in the transverse plane.

It is not yet clear why a concave curve is perceived as "straight" haptically, and this question will be investigated in Exp. III. It is clear from Exp. I, however, that blind *S*s (*a*) are less influenced by the effect, whatever its nature, and (*b*) show some important differences in scanning strategies, compared to the sighted *S*s we observed. Experiment I provides only a correlational link between categorization performance and scanning. The

TABLE 2

MEAN NUMBER OF INCORRECT CATEGORIZATIONS AS A FUNCTION OF SCANNING METHOD

| Scanning method | $\bar{X}$ | *SD* |
|---|---|---|
| Grip (pooled) | 11.95 | 2.88 |
| Grip with thumb | 12.70 | 2.83 |
| Grip without thumb | 11.20 | 4.85 |
| Pinch (pooled) | 14.45 | 2.87 |
| One-finger pinch | 15.30 | 3.38 |
| Four-finger pinch | 13.60 | 2.26 |
| Top sweep | 14.45 | 3.49 |

effects on categorizing curves of the exploratory methods most often used by blind and sighted *S*s in Exp. I were tested experimentally in Exp. II.

## EXPERIMENT II

### *Method*

*Subjects.*—The *S*s were 60 male and female undergraduate and graduate students. All but a few were right-handed.

*Apparatus and procedure.*—The task and stimulus series were the same as described for Exp. I. The only change involved the way *S*s scanned the stimuli. One group of 20 *S*s explored using only the top sweep, another 20 used only the pinch, and the remaining 20, only the grip. The *E* demonstrated several times the scanning method *S* was to use. Following the demonstration, *S* donned the blindfold and practiced the method on the 16-mm. warm-up stimulus until *E* was satisfied the scanning was correct.

*Definition of scanning methods.*—The top sweep consisted of moving index and middle fingers smoothly across the top edge of the stimulus. These two fingers were held straight out and together, parallel to the table, and *S*s felt the edge with only the fingertips.

The grip involved curling four fingers over the front edge of the stimulus and pulling back lightly against it while sweeping back and forth. Half of the *S*s in the Grip group used the thumb to feel the back edge of the stimulus as other fingers explored the front edge; the other half did not use the thumb, keeping it tucked under their palms. The effect on judgments of use of the thumb when gripping was tested because some *S*s in Exp. I used the thumb with the grip but many did not.

Experiment II also tested the effects on judgments of a one-finger, and a four-finger pinch. The scorers in Exp. I reported cases where the pinch occurred with more than one finger on the front edge of the stimulus and that the pinch became more difficult to differentiate from the grip as more fingers were used to pinch. The one-finger pinch consisted of holding the stimulus edge in a pincer between the thumb and index finger while sliding smoothly back and forth across it. The remaining three digits were tucked under the palm. The four-finger pinch was like the one-finger pinch except that the remaining three digits "trailed" along the front edge behind the thumb-index pincer, instead of being tucked under the palm. Half of the *S*s in the Pinch group used the one-finger pinch and the other half, the four-finger pinch.

### *Results*

Gripping the stimulus led to more objective curvature judgments than either pinching or top sweeping. As in Exp. I, incorrect categorizations for all groups were confined almost exclusively to concave curves.

Table 2 summarizes the results of Exp. II. The table shows that gripping with and without the thumb led to a statistically similar number of incorrect categorizations, $t\ (18) < 1.0$. Also, pinching with one and with four fingers led to a similar number of incorrect categorizations, $t\ (18) < 1.0$. Error scores were pooled across the two Grip groups and the two Pinch groups. Comparing these pooled scores with the incorrect categorizations associated with the top sweep, a striking finding was that the pinch and the top sweep led to an identical number of incorrect categorizations, but *S*s using the grip made fewer categorization errors. A single classification analysis of variance showed that the difference in mean categorization errors among the three scanning methods was significant, $F\ (2,\ 57) = 3.69,\ p < .05$. Subsequent *t* tests indicated that significantly fewer incorrect categorizations were associated with gripping than with either pinching, $t\ (38) = 2.78,\ p < .01$, or top sweeping, $t\ (38) = 2.46,\ p < .02$.

## EXPERIMENT III

The results of Exp. I and II establish an experimental link between objective judgments of concave curves and the use of a particular kind of exploratory touch, called gripping. Two important questions about the role of scanning in haptic curvature judgments still remain: What information does gripping give about concave curves that cannot be gathered with the pinch, top sweep, or with other methods of exploratory touch? Second, why is the grip so singularly effective in gathering this information?

Judging curvature haptically may be a task where critical aspects of the stimulus can be obscured by the way it is explored. A concave curve describes an arc in the same (transverse) plane and about the same shape as the path of *S*'s arm sweep while scanning. The *S* may have difficulty in distinguishing between this sort of arm movement and stimulus shape. Exploring only the top edge of the concave curve, as with the pinch or top sweep, seems to maximize the stimulus-arc-arm-arc correspondence, perhaps masking information about stimulus curvature and reducing the objectivity of judgments.

If errors in categorizing such curves do stem from this masking effect, reducing the match between stimulus arc and arm movement should lead to greater objectivity. In Exp. III, we tested several ways of reducing this match. One procedure was to have *S*s judge curvature stimuli that describe arcs completely contained in the frontoparallel plane—arcs higher or lower in the center than the ends, with the chords parallel to the table. The arm movement no longer overlaps these arcs because they lie in a different plane and judgments should be more objective.

A second procedure tested was to "straighten out" the curved path followed by the fingertips. The arc described when the arm pivots from the shoulder (called whole-arm movement) is flatter than when the pivot is the elbow (called forearm movement). In the present study, the whole-arm arc is less like the concave stimulus arc and should lead to more objective judgments of the curves used in Exp. I and II (called horizontal curves). However, judgments of the stimuli describing arcs in the frontoparallel plane (called vertical curves) should be unaffected.

### *Method*

*Subjects.*—Sixty undergraduate and graduate students participated in the experiment. Nearly all *S*s were right-handed.

*Apparatus.*—The study called for a series of curvature stimuli arced in the frontoparallel plane in addition to the stimuli used in Exp. I and II (horizontal curves). The two types of curves, compared in Fig. 6, were identical in length and arc height, differing only in the orientation of the curvature. Eight symmetrical curves—four concave and four convex arcs—and a straight edge (duplicating the values of the horizontal curvature series) were cut from 1.5-mm. thick posterboard. Concavity in a vertical curve meant that the midpoint of the curve was higher than the ends, and convexity, the midpoint lower than the ends. The stimuli were held upright in the frontoparallel plane in a plywood holder as shown in Fig. 6.

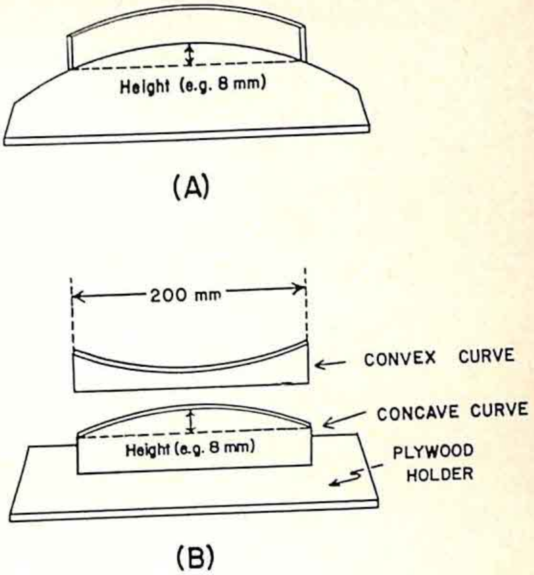


FIG. 6. A comparison of horizontal and vertical curves. (The upper figure, A, shows a horizontal concave curve like those used in Exp. I and II. The lower figure, B, shows a concave and a convex vertical curve.)

*Procedure.*—The task was the same one used in Exp. I and II. All *S*s scanned the stimuli with the preferred hand, using the top sweep described in Exp. II. (This method seems the best example of the overlap of arm movement with the arcs of horizontal concave curves.)

Forty *S*s judged the vertical curves and 20 *S*s, the horizontal curves. Half of the Vertical Curve group scanned with whole-arm movement and the other half, with forearm movement. The Horizontal Curve group *S*s used only whole-arm movement. (A Forearm Movement group to judge horizontal curves was unnecessary. The Top Sweep group in Exp. II scanned the horizontal curves with what has been defined as forearm movement and the data from that group were used again in Exp. III.) The *S*s using forearm movement kept their elbows in the elbow rest, whereas *S*s using whole-arm movement made all movements from the shoulder and kept their elbows "locked."

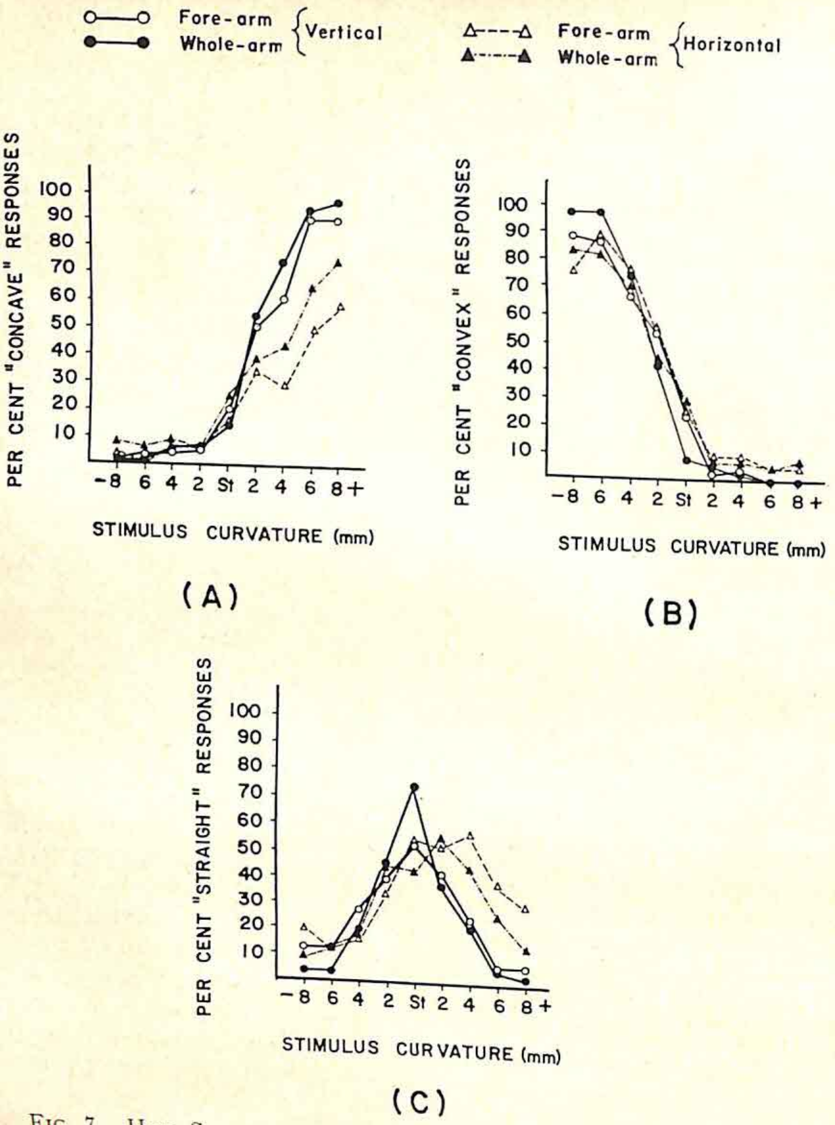


FIG. 7. How *S*s categorized stimuli in Exp. III. (The figure shows the frequency of "concave," A, "convex," B, and "straight," C, responses by each group of *S*s to each curvature value. Concave stimulus values are represented on the abscissa as "+," convex values as "−.")

### *Results*

Categorizations of vertical curves were extremely objective in sharp contrast to the judgments of *S*s exploring horizontal curves. Again, the psychophysical scale for horizontal curvature was skewed toward concavity, and incorrect categorizations of these curves were confined to concave stimuli. Incorrect categorizations of vertical curves, however, were evenly distributed among the three response cate-

TABLE 3

MEAN NUMBER OF INCORRECT CATEGORIZATIONS BY *S*S JUDGING HORIZONTAL AND VERTICAL CURVES

| Type of arm movement | Curvature orientation | | | |
|---|---|---|---|---|
| | Horizontal curves | | Vertical curves | |
| | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* |
| Forearm | 14.45 | 3.94 | 9.50 | 5.55 |
| Whole arm | 13.80 | 5.43 | 7.55 | 3.52 |

TABLE 4

Psychophysical Scales of Curvature for Horizontal and Vertical Curves

| Group | Curvature thresholds (mm.) | | | | IU (mm.) | PSE (mm.) (phenomenal straightness) |
|---|---|---|---|---|---|---|
| | Convex (−) | | Concave (+) | | | |
| | AL | *SD* | AL | *SD* | | |
| Horizontal curves | | | | | | |
| Forearm movement | −1.57 | 5.30 | +7.55 | 10.54 | 9.12 | +1.48 |
| Whole-Arm movement | −2.82 | 5.28 | +4.03 | 6.57 | 6.85 | + .23 |
| Vertical curves | | | | | | |
| Forearm movement | −4.07 | 8.23 | +2.14 | 3.85 | 6.21 | + .16 |
| Whole-Arm movement | − .33 | 5.86 | +1.45 | 3.35 | 1.77 | + .80 |

Note.—PSE = Point of subjective equality. IU = interval of uncertainty. AL = absolute threshold.

gories, and the psychophysical scale for vertical curvature showed no skew. Also, whole-arm movement reduced the skew toward concavity in the horizontal curvature scale relative to forearm movement.

The mean numbers of incorrect categorizations per group are compared in Table 3. A Curvature Orientation × Arm Movement analysis of variance showed that forearm and whole-arm movement led to a similar number of errors regardless of curvature orientation, $F\ (1, 76) = 2.81,\ p > .05$. However, *Ss* judging vertical curves were more objective than *Ss* judging horizontal curves regardless of arm movement, $F\ (1, 76) = 28.33,\ p < .01$. Indeed, judging the vertical curves led to about eight incorrect categorizations, nearly half the number made with the horizontal curves.

Figure 7 shows the frequencies of "concave" (Fig. 7A), "convex" (Fig. 7B), and "straight" (Fig. 7C) responses by each group of *Ss* to each curvature value. Figure 7A shows that *Ss* judging vertical curves gave more "concave" responses than Horizontal Curve group *Ss* to all of the concave curves. Figure 7C shows that when Horizontal Curve group *Ss* did not categorize the concave curves correctly, they were judging these stimuli to be "straight." Finally, Fig. 7B shows an extremely high frequency of "convex" responses to the convex stimuli, either horizontal or vertical, for all groups.

As in Exp. I, psychophysical scales for curvature were derived from group percent "concave" and "convex" response distributions (Fig. 7A and B) following a least-squares procedure (Guilford, 1954, p. 128). The findings are summarized in Table 4. The table shows that the scales for horizontal curvature are skewed toward concavity and that the skew is less with whole-arm movement than with forearm movement. Moreover, whole-arm movement led to less variability in the PSE for psychophysical straightness (as measured by the size of the IU) on the horizontal curvature scale, and in the AL for horizontal concavity. Finally, Table 4 indicates no large skew toward concavity in judgments of vertical curves.

## General Discussion

Apparently the relationship between active handling and veridical haptic form perception is not based on either the scanning alone or the stimulus alone. Indeed, there seems to be an interaction between the stimulus attribute being judged and the style of exploration. The present findings describe a particular interaction where arm movement may obscure stimulus curvature when the curvature coincides with the arc of the arm sweep. When the scanning does obscure the curves, one exploratory technique—the grip—was more effective than others.

The grip appears to have at least two distinguishing characteristics. First, it focused attention on a stimulus part informative of the shape attribute being judged, while avoiding irrelevant or misleading features. Gripping the curves focused attention on the front edge of the stimulus, an arc in a different plane than sweeping arm movement and probably more informative of the stimulus shape than the top edge. In contrast, exploring only the top edge of the curve, as with the pinch or top sweep, seems to maximize the correspondence between stimulus arc and arm arc.

A second characteristic of the grip was that it permitted a global, simultaneous apprehension of the stimulus curve. The distinguishing feature of a curve is probably an ends-to-middle relationship. Veridical perception of such a shape haptically might depend on how well the scanning apprehends this relationship. Gripping involved feeling most of the front edge of the stimulus at once, perhaps allowing *S*s to kinesthetically "triangulate" the ends-to-middle relationship by feeling three or more points along the edge simultaneously. This kind of patterning may have been particularly important with horizontal concave curves, possibly counteracting the obscuring effect of sweeping arm movement.

Some earlier studies have hinted that both attention focusing and patterning of stimulation may be general functions of haptic exploratory scanning (cf. Abravanel, 1968; Gibson, 1962; Hatwell, 1960; Zinchenko & Lomov, 1960). There is additional evidence that both functions may also characterize visual exploratory activity (cf. Mackworth & Morandi, 1967; Zinchenko & Lomov, 1960). These studies uniformly suggest that attention focusing in scanning is linked to isolating informative stimulus features. However, the nature of the patterning function of scanning is less well understood. One hypothesis is that simultaneous apprehension, in allowing *S* to pattern relationships or to compare one stimulus part with others, may help to organize successive parts of a form into a whole (Abravanel, 1968; Hatwell, 1959), thereby reducing the requirement for integrating successive inputs in memory (Davidson, 1972).

The improved scanning of blind *S*s in the present experiment points to a flexibility of the haptic perceptual system such that different forms of haptic attention can be learned. A few studies have linked improved haptic scanning to superior form judgments by the blind (cf. Hatwell, 1959, 1960). In addition, style of exploration seems to be one factor that differentiates experienced, proficient braille readers from those who are unskilled (Critchley, 1953). The experienced reader's exploratory scanning seems to focus attention on the informative parts of the braille word, skipping over redundant aspects (Kederis, Siems, & Haynes, 1965; Nolan & Kederis, 1969). Taken together with studies that indicate improvements in scanning activity during ontogenetic development (cf. Abravanel, 1968; Zinchenko & Ruzskaia, 1960a, 1960b), the data suggest that development of skilled exploratory activity may comprise one important type of perceptual learning.

An understanding of the ontogeny of exploratory activity as a form of perceptual learning may help to exploit the apparent flexibility of the haptic modality to improve the information-gathering skills of blind *S*s. A second result would be a better understanding of the development of perceptual skill in the normal individual.

## REFERENCES

Abravanel, E. The development of intersensory patterning with regard to selected spatial dimensions. *Monographs of the Society for Research in Child Development*, 1968, **33** (No. 2, Serial No. 118).

Critchley, M. Tactile thought with special reference to the blind. *Brain*, 1953, **76**, 19–35.

Davidson, P. W. Haptic perception of curvature by blind and sighted humans: A study of the role of exploratory activity in perception. Unpublished doctoral dissertation, George Washington University, 1970.

Davidson, P. W. The role of exploratory activity in haptic perception: Some issues, data, and hypotheses. *American Foundation for the Blind Research Bulletin*, 1972, **24**, 21–27.

Gibson, J. J. Observations on active touch. *Psychological Review*, 1962, **69**, 477–491.

Goodnow, J. J., Baum, B., & Davidson, P. W. A haptic error: Skew in a symmetrical curve. *Perception & Psychophysics*, 1971, **10**, 253–256.

Guilford, J. P. *Psychometric methods.* N. Y.: McGraw-Hill, 1954.

Hatwell, Y. Perception tactiles des formes et organisation spatiale tactile. *Journal de Psychologie*, 1959, **56**, 187–204.

Hatwell, Y. Etude de quelques illusions geometriques tactiles chez les avegules. *L'Annee Psychologique*, 1960, **60**, 11–27.

Hunter, I. M. L. Tactile-kinesthetic perception of straightness in blind and sighted humans. *Quarterly Journal of Experimental Psychology*, 1954, **6**, 149–154.

Kederis, C., Siems, J., & Haynes, R. A frequency count of the symbology of English braille grade 2.

American usage. *International Journal for the Education of the Blind*, 1965, **15**, 38–46.

Mackworth, N., & Morandi, A. The gaze selects informative details within pictures. *Perception & Psychophysics*, 1967, **2**, 547–552.

Nodine, C., & Lang, N. The development of visual scanning strategies for differentiating words. *Developmental Psychology*, 1971, **5**, 221–232.

Nolan, C. Y., & Kederis, C. *Perceptual factors in braille word recognition*. N. Y.: American Foundation for the Blind, 1969. (Research Series, No. 20)

Pick, A. D., & Pick, H. L. A developmental study of tactile discrimination in blind and sighted children and adults. *Psychonomic Science*, 1966, **6**, 367–368.

Rubin, E. Haptische untersuchungen. *Acta Psychologica*, 1936, **1**, 285–380.

Scott, W., & Wertheimer, M. *Introduction to psychological research*. N. Y.: Wiley, 1962.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. N. Y.: McGraw-Hill, 1956.

Zinchenko, V., & Lomov, B. The functions of hand and eye movements in the process of perception. *Problems of Psychology*, 1960, **1**, 12–26.

Zinchenko, V., & Ruzskaia, A. Comparative analysis of touch and vision: Report III: Visual-haptic transfer in preschool children. *Doklady Akademii Pedagogicheskikh Nauk RSFSR*, 1960, **3**, 95–98 (Joint Publications Research Services translation, Document No. 10686). (a)

Zinchenko, V., & Ruzskaia, A. Comparative analysis of touch and vision: Report IV: Development of tactile orientation and haptic-visual transfer in preschool children. *Doklady Akademii Pedagogicheskikh Nauk RSFSR*, 1960, **3**, 99–102 (Joint Publications Research Services translation, Document No. 10686). (b)

(Received June 21, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 56–62

# SOME CONTEXT EFFECTS IN THE SPEEDED COMPREHENSION OF SENTENCES [1]

D. JAMES DOOLING [2]

*Kent State University*

A target sentence was shown following the presentation of a sentence context, a one- or two-word context, or no context, which represented decreasing degrees of semantic content. The contexts were either appropriate or inappropriate to the target sentences. Time to comprehend the sentences was measured according to two different definitions of "comprehension." In Task 1, *S* had to judge whether a sentence was meaningful or anomalous. Reaction times for correct judgments with appropriate context were: sentence context < word context < no context. In Task 2, *S* had to judge if a target sentence was appropriate for the preceding context. Reaction times were: sentence context > word context. These results suggest that context affects different components of the sentence comprehension process in opposing ways. Context speeds up the processing of the basic semantic relationships of a sentence. But the presence of context requires an additional process: context-sentence integration.

When *S*s read connected discourse, they abstract out its theme and use it as a mnemonic device to improve their performance on a retention test (Dooling & Lachman, 1971; Pompi & Lachman, 1967). This suggests that it is typical and natural for sentences to be comprehended as part of a larger semantic unit. In natural language, there is a semantic structure among the sentence units and, therefore, the meaning communicated by an individual sentence is, to some degree, redundant with the context. This semantic structure and redundancy should be of some use in the comprehension of individual sentences. It is reasonable, therefore, to ask: Can *S*s use contextual meaning to "short circuit" some of the semantic processing involved in arriving at the meaning of an individual sentence? Does context speed up the sentence comprehension process?

An experiment was performed which attempted to answer this question in a straightforward manner (Dooling, 1971, Exp. I). Comprehension time for various sentences was measured under conditions of prior context versus no context. The measure of comprehension time was similar to that used by Danks (1969). When the sentence was presented, a clock started. The *S* then read the sentence and indicated when he understood it by pushing a button which stopped the clock. The results did not support the hypothesis. There was no significant overall effect of context. In fact, the context conditions led to slightly *longer* comprehension times. These findings led to a reevaluation of the experimental task and to a more complex hypothesis as to the effects of context. With the subjective measure of comprehension that was employed, *E* had no control over the criterion that *S* set for himself in making a response. Under these conditions, *S* may have used different subjective definitions of "comprehension" in the context and no-context conditions. With no prior context, *S* would compute the basic sentence meaning and respond. But the presence of context would affect the time to respond in two opposing ways. Context may speed up the comprehension of the basic meaning of the sentence, but the presence of context may add a process that requires time to perform: the sentence meaning is integrated with the context. Sentences which are understood in context are, in some sense, given a "deeper" meaning.

---

[1] This study was based on a doctoral dissertation submitted to the graduate school of the State University of New York at Buffalo. I would like to acknowledge the contributions of the following members of my dissertation committee: Roy Lachman, Erwin Segal, and Irving Biederman. I am grateful as well to the following for their helpful comments on an earlier draft of this paper: Stephen F. Checkosky, Joseph H. Danks, Joseph S. Dumas, Roy Lachman, Janet L. Mistler-Lachman, and Erwin Segal. A shorter version of this paper was read at the Midwestern Psychological Association convention in Cincinnati, Ohio, on April 30, 1970.

[2] Requests for reprints should be sent to D. James Dooling, Department of Psychology, Kent State University, Kent, Ohio 44240.

The present experiment tests the above interpretation of context-sentence processing. The strategy involves using two experimental tasks that actually represent two different definitions of "comprehension." In Task 1, *S* has to make a speeded judgment as to whether a target sentence is meaningful or anomalous. To do this, he must compute the basic semantic relationships of the sentence, but no context-sentence integration is required. Task 2, on the other hand, requires a rather explicit process of context-sentence integration. The *S* must make a speeded judgment as to whether or not the target sentence follows appropriately from the context. The same contexts and sentences are used for both tasks; it is solely the nature of the speeded decision that is varied between them. In each case, the time for a correct YES response is taken as the estimate of processing time. The NO trials are included as foils to keep *S* honest, but are not of direct interest in themselves.

Unlike the previously mentioned study, the major variable is not context versus no context, but rather two degrees of context. It is, of course, impossible to have a no-context condition for Task 2, and it is desirable to compare comprehension times on the two tasks using the same contexts and sentences. But more importantly, it is not legitimate to make direct comparisons of processing times from two different tasks. The time for a correct YES on each task will be affected not only by the comprehension processes required, but also by the difficulty of the YES–NO discrimination. We do not know if the anomalous sentences chosen as foils for Task 1 are equal in difficulty with the inappropriate contexts employed in Task 2. For this reason, two levels of context are employed: word context and sentence context. The sentence contexts provide a greater amount of prior semantic information. In Task 1, a sentence context should result in faster reaction times. But in Task 2, the sentence contexts will require that more information be integrated with the basic meaning of the target sentence. The facilitative effects of the longer context will be offset by the fact that there is more context to integrate. The major prediction, therefore, is a Context × Task interaction on the correct YES reaction times. In addition to the two levels of appropriate context, as described above, no-context and inappropriate-context conditions are also included for Task 1. This allows us to evaluate context effects across a wider range for the particular level of comprehension required by Task 1.

The experiment also includes a test of a subsidiary hypothesis. Fodor, Garrett, and Bever (1968) have argued that the main verb of a sentence is a central element in sentence comprehension. Similarly, Healy and Miller (1970, 1971) have presented evidence that the verb is more influential than the subject in determining the meaning of a sentence. Segal and Martin (1966), on the other hand, had *S*s rate the importance of words in sentences and found the subject to be most important. In the present experiment, there are conditions where either the subject or main verb is given ahead of time as context. Which will have the greater effect on Task 1 comprehension speed? If lexical analysis of the main verb has a primary role in sentence comprehension, then advance knowledge of the verb should yield the faster times. The experiment also includes a stronger test of the Fodor et al. (1968) hypothesis. In some cases, both subject and object are included in the context. If the verb is, in fact, the primary operator in sentence comprehension, its prior presentation may facilitate comprehension speed as much as the priming of two other words from the sentence. When context is presented, therefore, it will focus on one of these three sentence elements: (*a*) subject, (*b*) main verb, (*c*) subject plus another word, usually the direct object.

## METHOD

*Subjects.*—The *S*s were 6 male and 6 female students of introductory psychology who were fulfilling a course requirement at the State University of New York at Buffalo. Because the experiment took longer than the required 3 hr., they were also paid $3.00 for their services.

*Design.*—All experimental variables were within *S*. They were as follows. (*a*) Task: Task 1 required *S* to make a YES–NO reaction time judgment as to whether or not the target sentence was meaningful. Task 2 required a YES–NO reaction time judgment as to whether or not the target sentence followed appropriately from the context. (*b*) Context: Word contexts consisted of one or two words from the target sentence. Sentence contexts consisted of a sentence which contained one or two words from the target sentence. For Task 1 only, there was a condition in which no prior context was presented. (*c*) Linguistic focus: Context highlighted one of these sentence elements—the subject, main verb, or subject plus another word. (*d*) Sentence type: Sentences of four different grammatical structures were employed.

*Sentences.*—All of the target sentences to be comprehended were five words long and conformed to one of the four grammatical structures as follows (comprehension times are included in seconds):

1. The N V the N.
   The clown amused the children. (.84)
2. Adj N V Adj N.
   Pink bouquets emit fragrant odors. (1.10)
3. The Adj N V Adv.
   The old man walked slowly. (.98)
4. PropN V Prep the N.
   John played in the yard. (.89)

In these four different sentence frames, the subject and the verb can occupy three different serial positions. This feature of the sentences discourages any special scanning strategy when one of these elements is mentioned in the context. A total of 236 sentences was employed in the study, 59 of each type. For each of the meaningful sentences, three anomalous counterparts were constructed. In each, a semantic rule was violated, but syntax remained intact. The anomalous sentences always contained the word or words highlighted by the context. This feature forced *S* to read the sentence before making his speeded judgment. The definition of meaningful versus anomalous was validated by three raters. Examples of the anomalous sentences are given in the bottom portion of Table 1. The complete set can be found in Dooling (1971).

*Contexts.*—For each target sentence, 12 contexts were constructed as required by orthogonal combination of the three relevant variables: appropriateness, amount of semantic information, and linguistic focus. An appropriate word context consisted of either one or two words from the target sentence, as defined by the linguistic focus variable. Inappropriate contexts were obtained by mismatching the appropriate ones. The appropriate sentence contexts consisted of the word or words from the word contexts along with added semantic information. Each was a short descriptive sentence that could serve as a meaningful lead-in to the target sentence. About half of the inappropriate sentential contexts contained appropriate lexical items. (The examples in Table 1 are of this type.) The other half were mismatches of appropriate contexts. Because the inappropriate sentential contexts sometimes contained words repeated in the target sentence, *S* had to read each sentence and could not make a decision based only on the occurrence of one or two key words. The appropriateness of the contexts was verified by the ratings of three college students. Table 1 gives examples of the different types of context for one of the target sentences, "The baby vomited the milk." The complete set of 2,832 contexts can be found in Dooling (1971).

*Procedure.*—The *S*s were told that the experiment was designed to find out how context influences the speed with which we understand sentences. It was

TABLE 1

EXAMPLES OF CONTEXTS AND ANOMALOUS SENTENCES FOR THE TARGET SENTENCE: "THE BABY VOMITED THE MILK."

| Sentence type | Linguistic focus | Word | Sentence |
|---|---|---|---|
| Context | | | |
| Appropriate | Subject | baby | A baby was given a sour bottle before her nap. |
| | Verb | vomited | Vomiting occurred when a sour bottle was given before the nap. |
| | Subject plus | baby milk | A baby was given sour milk before her nap. |
| Inappropriate | Subject | storm | A baby was kidnapped from her carriage. |
| | Verb | approached | Vomiting occurred when the relatives ate the spoiled potato salad. |
| | Subject plus | storm rapidly | The baby became healthier every day due to the nourishing milk. |
| Anomalous | Subject | | The baby drank the house. |
| | Verb | | The desk vomited the shoe. |
| | Subject plus | | The baby spoke the milk. |

emphasized that "understanding" referred to the normal, effortless process of reading for meaning. The *S*s were fully informed about the number of trials and blocks and the fact that two-thirds of the correct responses were "yes." They were encouraged to respond as fast as they could, but to keep their error rate at about 10%.

The instructions for Task 1 told *S* to respond YES if the target string of words made up a meaningful English sentence and to respond NO if it did not. It was emphasized that context would be relevant to the target sentence on most trials. For Task 2, *S* was told to respond YES if the target sentence was meaningful and appropriate for the context, and to respond NO if it was not. The *S*s were told about the two types of trials requiring a NO response and were cautioned not to adopt a specialized strategy for the experiment.

Each trial began when *S* pressed an advance key, revealing the context in the top portion of a one-way window. Following a second button press, the window was blank and *S* repeated the context aloud. This procedure ensured that the context had been registered properly and was "on *S*'s mind" before presentation of the target sentence. A third press on the advance key presented the target sentence and started a clock. The clock stopped when either a YES or NO button was pressed. At that point, *E* called out the time in hundredths of a second, also informing *S* if he had made an error. The intertrial interval was determined by the time needed by *E* to reset the apparatus and change the stimulus materials, about 2 or 3 sec.

All *S*s were run in all conditions of the experimental design on each of 2 days. Day 1 was considered practice. All conditions for a given task were run together in a single session lasting approximately 45 min. Two such sessions were run on each day, separated by a 10-min. break. The six male *S*s did the tasks in the order 1-2-2-1 across the 2 days. For the females, it was Task 2-1-1-2. There were seven blocks of trials for each session of Task 1, and six for Task 2. (The extra block for Task 1 was for the no-context condition). A given block was homogeneous with respect to the context (word vs. sentence vs. none) and linguistic focus (subject vs. verb vs. subject-plus) variables. For example, in one of the blocks of trials, all contexts might be in sentence form and focus on the subject. Because of this blocking, *S* knew what form of context to expect and did not have to switch his set from trial to trial. Each block consisted of 18 trials, 12 YES and 6 NO. In Task 1, 8 of the YES responses were to sentences which followed appropriate contexts, and 4 were not. In Task 2, 3 of the NO trials involved an inappropriate relationship between the context and a meaningful sentence. On the other 3 NO trials, the target sentence itself was anomalous, but contained one or more words from the context. The presence of these types of NOs prevented *S* from making his Task 2 judgment by merely scanning the target sentence for a key word. The blocks were counterbalanced within each task and the 18 trials within a block were randomly ordered. The target sentences were assigned to conditions by a computer algorithm, sampling from the 236 possibilities randomly and without replacement. Assignment of a target sentence to 1 of the 234 (18 Trials × 13 Blocks) experimental conditions determined which context would be selected. Twenty-four sets of materials were generated in this manner—a set for each *S* on each day.

## RESULTS AND DISCUSSION

The dependent measure of major interest was the time for a correct YES reaction time on the second day of practice. The major analysis consisted of a repeated-measures analysis of variance with the variables: (*a*) task (sentence meaning vs. context-sentence appropriateness), (*b*) context (word vs. sentence), (*c*) linguistic focus (subject vs. verb vs. subject-plus), and (*d*) sentence type (four different sentence frames). Context was appropriate to the target sentence on all trials considered in this analysis. Two main effects were significant: linguistic focus, $F(2, 22) = 8.6$, $p < .01$, and sentence type, $F(3, 33) = 38.2$, $p < .001$. Neither context nor task was a significant main effect. There was one significant interaction, that of Task × Context, $F(1, 11) = 12.6$, $p < .01$. These significant effects will be described and discussed under three separate headings. Then the results of two secondary analyses will be reported.

*Context × Task interaction.*—The significant interaction reflects the fact that context had opposite effects on comprehension time in the two experimental tasks. The means for this effect are shown in Table 2. In Task 1, comprehension times decrease as context increases; the two types of context yielded a difference of .05 sec. The more context, the faster *S* could compute the basic semantic relationships of the sentence. But context affected performance in Task 2 in quite the opposite way. When required to match the sentence meaning to the context, longer contexts meant longer comprehension times (a difference of .06 sec.). This predicted interaction was in no way compromised by a higher order interaction. In fact, the direction of these effects was the same for all three conditions of linguistic focus and all

TABLE 2

Mean Hundredths of a Second for a Correct Yes Reaction Time as a Function of Task and Context

| Task | Context | |
|---|---|---|
| | Word | Sentence |
| 1 | 98 (.03) | 93 (.01) |
| 2 | 92 (.04) | 98 (.05) |

Note.—Mean proposition of false negative errors are included in parentheses.

four sentence types. The error data, also shown in Table 2, reflect the same pattern. The reaction time results, therefore, cannot be attributed to a speed-accuracy trade off.

The Task 1 results show that prior semantic information can be used to speed up the sentence comprehension process. Evidently, the skilled reader has the ability to use contextual information to short circuit part of the semantic processing in computing the meaning for a sentence. Although we do not understand what these processes might be, the contextual effect does bear some similarity to what Neisser (1967) has called the "word apprehension effect." Just as letters are not individually recognized in the perception of words, so too, sentences in context may not need full semantic analysis. But in assessing context effects, a second point must not be overlooked. The Task 2 results show that additional context can slow down part of the comprehension process. Comprehending a sentence in context is a more complex task than comprehending a sentence in isolation. It requires more cognitive work to relate the meaning of a sentence to its context. The experimental analogue used to demonstrate this point may need further elaboration in order to deal with two ways in which context could add to the processing load. (*a*) Sentences understood in context may be given a "deeper" meaning, that is, a semantic analysis that is highly specific to the context. (*b*) In comprehending sentences in discourse, *S*s construct a "theme" or "schema"; they recode the information into larger semantic units, and this process takes time to perform.

According to the rationale of the present study, Task 2 performance represents an addition to the processing required in Task 1. Other things being equal, we would expect comprehension times in Task 2 to be longer than for Task 1. In the present results, however, the overall times for the two tasks were nearly identical. The following arguments suggest that other things were not in fact equal and that the results are not seriously compromised by this apparent anomaly. (*a*) Direct comparison across the two tasks is not legitimate because times for a YES response will vary with the difficulty of the YES–NO discrimination. It was this logic which led to the inclusion of two levels of context for the experiment. (*b*) If sentences are normally understood by integrating their meaning with the context, then Task 1 requires an unnatural process: *S* must make a premature "exit" from his normal sentence comprehension routine. This process may have artificially inflated the reaction times in Task 1. (*c*) The error rates for Task 2 were twice as high as those for Task 1.

*Linguistic focus.*—Correct YES reaction times were significantly affected by the sentence elements that were highlighted by the context. The means were: subject, .96 sec.; verb, .98 sec.; and subject-plus, .92 sec. No prediction was made about this main effect, although it is surprising that the generally longer subject-plus-another-word contexts led to the fastest times in both experimental tasks. Contexts which contained two sentence elements might have been expected to yield longer Task 2 reaction times.

The linguistic focus variable was included in the study primarily to test a subsidiary hypothesis about the relative importance of the subject and verb in sentence processing. Since the hypothesis concerns the interaction of linguistic structures in computing the meaning of a sentence, only Task 1 performance is relevant to the issue. Also limiting the analysis to word contexts, we find that prior presentation of the subject or verb of a sentence yielded identical mean comprehension times of 1.01 sec. Two-word contexts (usually subject plus object) led to faster Task 1 performance (.93 sec.). The results seem to indicate that neither subject nor verb is central for comprehension. There is, however, another possibility. In postexperimental interviews, several *S*s felt that prior presentation of the subject was more helpful. The reason given was that it came up closer to the beginning of the sentence. In the case of the verb, they always had to hold it in memory while reading the subject. This points out a confounding that makes the present experiment an inappropriate test of the hypothesis in question. The verb always came later than the subject in the sentences. If intervening words do interfere, then the present results might be taken as indirect support for the verb-as-operator hypothesis. The question must be addressed, however, in a new experiment which systematically varies the serial position of the subject and verb.

*Sentence type.*—Target sentences of four grammatical structures were employed for the sake of generality. The mean times for the individual sentences were given in the Method section under sentences. It is of some interest to note that the type which took the longest to comprehend also has the most complex deep structure. It is also evident that the reaction times are closely related to the number of content words. These results differ from the findings of Fodor and Garrett (1967). They found that adding three adjectives did not affect the comprehension of sentences with two self-embedded clauses. The difference may lie in the fact that the Fodor and Garrett sentences are not "natural" for *S*s and may not be comprehended in the normal manner (cf. Marks, 1968).

*No context.*—The analysis described above did not include the condition in which no verbal context preceded the target sentence. The Task 1 results showed that more context meant faster comprehension. But that does not necessarily mean that the presence of context affects speed of comprehension. More direct evidence can be obtained on this point by comparing context versus no-context conditions, using Task 1 performance as our definition of comprehension. The means are shown on the right-hand side of Table 3 for zero, word, and sentence contexts, respectively. (The two columns at the left present means with inappropriate contexts, to be discussed below.) Context appears to have a sizable influence on comprehension speed. This was confirmed by analysis of variance; the main effect of context was highly significant, $F$ (2, 20) = 16.4, $p < .001$. The percentage of errors reflects the same pattern of results.

TABLE 3

MEAN HUNDREDTHS OF A SECOND FOR A CORRECT YES REACTION TIME AND MEAN PROPORTION OF FALSE NEGATIVE ERRORS

| Measure | Context | | | | |
|---|---|---|---|---|---|
| | Inappropriate | | | Appropriate | |
| | Sentence | Word | Zero | Word | Sentence |
| Comprehension time | 113 | 123 | 114 | 98 | 93 |
| Errors | .12 | .12 | .04 | .03 | .01 |

*Inappropriate contexts.*—A logical extension of the above results (that a good context helps) is that a misleading context will slow down processing. One-third of the YES trials in Task 1 were performed under conditions where the context was inappropriate. This allows us to compare the two levels of inappropriate contexts with zero context. The means are shown on the left-hand side of Table 3. In an analysis of variance paralleling that reported above, the effect of context was significant, $F$ (2, 20) = 4.35, $p < .05$. But only the word contexts led to worse reaction time performance than zero context. The only performance decrement with inappropriate sentence contexts was with the error rates. It seems clear, nevertheless, that an inappropriate context interferes with sentence comprehension. It was not predicted that a misleading word would have a greater negative effect than a misleading sentence. No explanation for this anomaly is presently available.

## REFERENCES

Danks, J. H. Grammaticalness and meaningfulness in the comprehension of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 678–696.

Dooling, D. J. *Some context effects in the speeded comprehension of sentences* (Doctoral dissertation, SUNY at Buffalo) Ann Arbor, Mich.: University Microfilms, 1971, No. 71-6065.

Dooling, D. J., & Lachman, R. Effect of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, **88**, 216–222.

Fodor, J. A., & Garrett, M. Some syntactic determinants of sentential complexity. *Perception & Psychophysics*, 1967, **2**, 289–296.

Fodor, J. A., Garrett, M., & Bever, T. G. Some syntactic determinants of sentential complexity: II. Verb structure. *Perception & Psychophysics*, 1968, **3**, 453–561.

Healy, A. F., & Miller, G. A. The verb as the main determinant of sentence meaning. *Psychonomic Science*, 1970, **20**, 372.

Healy, A. F., & Miller, G. A. The relative contribution of nouns and verbs to sentence acceptability and comprehensibility. *Psychonomic Science*, 1971, **24**, 94–96.

Marks, L. E. Scaling of grammaticalness of self-embedded English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 965–966.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Pompi, K. F., & Lachman, R. Surrogate processes in the short-term retention of connected discourse. *Journal of Experimental Psychology*, 1967, **75**, 143–150.

Segal, E. M., & Martin, D. R. The influence of transformational history on the importance of words in sentences. Paper presented at the Psychonomics Society Convention, St. Louis, October 1966.

(Received July 1, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 63–68

# DESCRIPTIVE VERSUS NORMATIVE MODELS OF SEQUENTIAL INFERENCE JUDGMENT[1]

JAMES SHANTEAU[2]

*University of Michigan*

Sequences of pink and white beads were drawn with replacement from one of two boxes. After each bead was displayed, *S*s made inference judgments by estimating the probability that the beads were drawn from the box with more white beads. Stimulus sequences were constructed from factorial designs which permitted simple tests of additivity as well as evaluation of serial-position and diagnosticity effects. The judgments were additive in probability form but less additive in Bayesian log-odds form and contained both general recency effects and small diagnosticity effects. A normative Bayesian model did not do as well with these results as a descriptive model from information-integration theory. Both models were able to handle diagnosticity effects, but only the descriptive model could handle additivity and serial-position effects.

Inference judgments have been most frequently studied in one experimental situation: the "book-bags-and-poker-chips" task (Edwards, 1968; Slovic & Lichtenstein, 1971). In this task, a sample of chips is drawn from one of two bags of poker chips, and *S* makes a probability judgment about which bag was sampled.

Such judgments have been studied almost exclusively within a normative, or Bayesian framework. Despite such shortcomings as "conservatism," the Bayesian formulation has not been seriously challenged as a model for inference behavior. Recent evidence suggests, however, that a model derived from information-integration theory might provide an attractive alternative (Shanteau, 1970). The purpose of this paper is to compare the integration and Bayesian models.

The Bayesian model is based on an optimal statistical rule which states that the final (posterior) opinion is a product of the initial (prior) opinion and the (likelihood) ratio of the probabilities that each bag produced the sample. When the two bags contain symmetric proportions of chips of two colors, e.g., pink and white, this rule can be written as

$$\frac{R_n}{1 - R_n} = \left(\frac{p}{1 - p}\right)^{W-P} \frac{R_0}{1 - R_0}. \quad [1]$$

In words, the probability after $n$ chips are sampled, $R_n$, is a function of (*a*) the population proportion $p$, (*b*) the difference $W - P$ in the number of sampled chips of each color, and (*c*) the initial opinion $R_0$.

When inference judgments are compared to this model, they are typically "conservative," or less extreme than predicted. This has been handled by modifying the likelihood ratio by an "accuracy ratio." Otherwise, the judgments do appear to be a function of $W - P$, $R_0$ and, to a lesser extent, $p$ (Edwards, 1968). This led previous investigators to claim that the Bayesian model provides a basis for understanding inference behavior.

However, Shanteau (1970) has suggested that the Bayesian interpretation of the book-bags task may be inappropriate. In an initial experiment, he found that *inference* judgments were indistinguishable from *estimates* of population proportion. This was quite unexpected since logically inference and estimation instructions should have produced quite different behavior.

---

[1] This investigation was completed during the author's tenure in United States Public Health Service Postdoctoral Fellowship 1 FO2 MH-36, 963-01 from the National Institute of Mental Health and was supported in part by National Science Foundation Grant GB-6666. The author wishes to thank Norman H. Anderson for his advice during all stages of work. The author also wishes to thank Brown Grier and Thomas Wallsten for helpful comments on the manuscript, and Deborah Spencer for her assistance in running *S*s.

[2] Requests for reprints should be sent to James Shanteau, who is now at the Department of Psychology, Kansas State University, Manhattan, Kansas 66502.

The same result was found in a second experiment which included special instructional checks and longer stimulus sequences. This suggests that *S*s may not be making inferences at all, but may instead be making estimation judgments. However, Shanteau did not numerically specify the population proportions, and so the results are not critical for the book-bags task.

Beach, Wise, and Barclay (1970) also criticized the book-bags task after finding that inferences were closely tied to sample proportions. Unfortunately, they failed to use a control group with estimation instructions; this was a crucial feature of Shanteau's (1970) experiment.

As an alternative to the Bayesian model, Shanteau (1970) proposed that Anderson's (1971) integration theory might better account for inference judgments. This descriptive approach assumes that each sample has both a scale value and a weight or importance value. The response after $n$ samples is then a weighted average

$$R_n = \sum_{i=0}^{n} w_i s_i, \qquad [2]$$

where the weights sum to 1. For the book-bags task, the scale values are 0 or 1, depending on the color of the sampled bead, and the weight values are assumed to vary with the serial position of the sample. In the present application, this implies that the judgments should be additive. Functional measurement techniques may then be used to test the model and to evaluate its parameters (Anderson, 1971).

The two models can be compared on three grounds. First, the Bayesian model predicts that the diagnosticity, or importance, of a sample should vary according to the population proportion $p$. The integration model, in contrast, allows diagnosticity to depend on any factor which influences the initial weight $w_0$. Second, the Bayesian model predicts that the responses should be additive after a log-odds transformation; this can be seen by taking the logarithm of Equation 1. In comparison, the integration model assumes that the responses are additive under a monotonic transformation (Anderson, 1971). Since the raw responses are typically appropriate (Anderson, 1971) and Shanteau (1970) found that probability responses were additive, no transformation was applied to the present judgments. Finally, the integration model allows serial position effects to play a significant part in the judgments, while the Bayesian model predicts zero (no) order effects.

## METHOD

Sequences of pink and white beads were drawn with replacement from one of two boxes. After each bead was displayed, *S* estimated the probability that the beads were drawn from the box with more white beads. Two levels of diagnosticity were used with different *S*s. In the 60/40 condition, one box contained 60 white and 40 pink beads and the other contained 40 white and 60 pink beads. The 75/25 condition used boxes with contents of 75 and 25.

*Stimulus sequences.*—All *S*s saw the same 24 sequences of 4 to 11 beads. The middle 18 sequences were constructed from two factorial designs. A $2^4$ factorial specified 16 sequences so that the first 4 serial positions contained all possible combinations of pink and white beads. A second factorial design specified the sequences over 9 serial positions. Since a complete $2^9$ factorial was impractical, a highly confounded, orthogonal main-effects design was used (Addelman & Kempthorne, 1961, p. 140). These sequences were arranged to contain (overlap) most of the $2^4$ sequences over the first 4 serial positions. This allowed complete tests of additivity for the fourth responses and confounded tests for the ninth responses. This also allowed derivation of serial position weights for both response positions.

Two precautions were taken in arranging the sequences. First, the sequences were presented in six different orders according to a Latin square arrangement. This permitted checks of possible order-of-sequence effects; however, no significant effects were observed in either the 60/40 or 75/25 condition. Second, the first and last three sequences were identical to check on the consistency of the judgments over the experiment; however, the responses turned out to be relatively invariant.

*Procedure.*—For practice, *S* initially drew a sequence of beads from each of the boxes. After each draw, *S* used the accumulated evidence to estimate the probability that the beads came from the mostly white box. After replacing the bead, the box was mixed up and another bead drawn. Following this practice, *S* was told that to save time, *E* would draw the remaining beads.

Each sequence began by *E* flipping a coin out of *S*'s view to determine which box to use. The selected box was placed in front of *S* and beads drawn from it. While this procedure was used to convince *S* that the draws were random, a special third box was actually used. This box was partitioned into one part with only white beads and

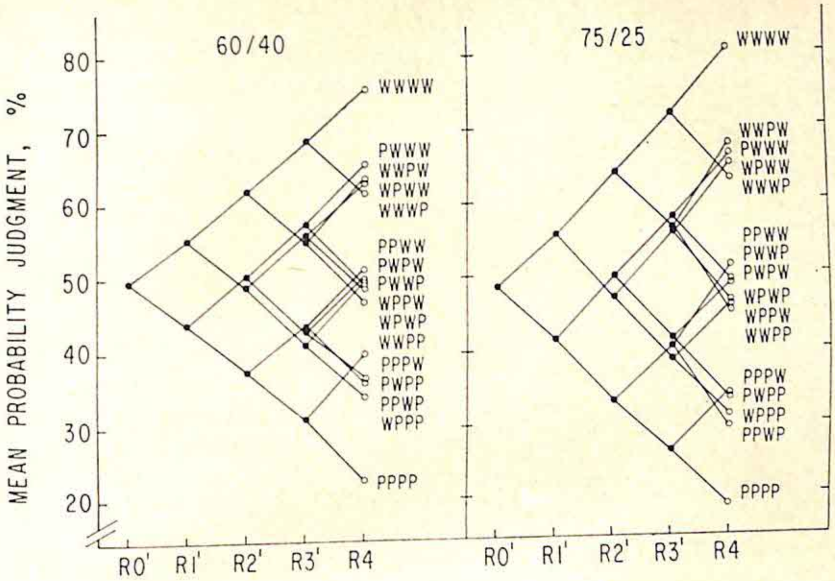


FIG. 1. Backward projection of R4 judgments for 60/40 and 75/25 conditions: additivity analysis for integration model.

another with only pink beads; this allowed *E* to vigorously mix up beads while drawing a bead of the prescribed color. After each sequence, the box was placed behind a partition and *E* flipped the coin to select the next box. The entire session lasted 50 min.

All responses were made on 14, 4-in. lines displaced horizontally on a $8\frac{1}{2}\times 11$ in. sheet. As each bead was drawn, *S* recorded its color as "W" or "P" on the left edge of the page. The probability estimates were then made by making a pencil mark through the line. After each estimate, *S* covered it and all previous estimates with a piece of cardboard; this kept *S* from using his previous marks as an artificial memory aid. To correspond to probabilities, these ratings were recorded to the nearest .04 in. on a 0 to 100 scale.

Each line was described as representing a range of probabilities from 0 to 1. The *S* was told to mark the right end if he was absolutely certain the mostly white box was picked. If he was totally uncertain, he was to mark the middle. If he was absolutely certain the mostly pink box was picked, then he was to mark the left end. To check *S*'s understanding, he was initially required to correctly mark the probabilities .25 and .90 on a line.

Two procedural precautions were taken in presenting the stimuli. First, filler beads were added to the ends of most sequences to avoid end-of-sequence response effects. Second, the contents of the boxes were visually defined by filling clearly labeled plastic bags in exactly the same way as the boxes. These constantly visible bags were expected to enhance any diagnosticity effects by making the box proportions more concrete.

*Subjects.*—Forty-eight undergraduates fulfilling a class requirement were randomly assigned in equal numbers to the two diagnosticity conditions; six additional *S*s were eliminated due to procedural difficulties. Pairs of *S*s were run together, but separated by a partition.

## RESULTS

The fourth (R4) and ninth (R9) serial judgments were selected for detailed examination. The mean R4 values appear as open circles in Fig. 1. The actual bead sequences are labeled next to the points, with *W* and *P* indicating white beads and pink beads, respectively.

*Diagnosticity.*—The R4 values in two panels of Fig. 1 reveal a very similar pattern with the 75/25 judgments more extreme by an average of 2.7%. However, this difference was not significant as shown by the Condition × Sequence interaction. For R9 judgments, the 75/25 values were more extreme by an average of 4.3%, which was significant, $F$ (11, 506) = 2.25.

This finding suggests that the diagnosticity effect of population proportion increases across serial position. Since previous workers have typically collapsed their results across serial position, this may account for why some investigators report a diagnosticity effect (e.g., Edwards, 1968), while others report little or no effect (e.g., Pitz, Downing, & Reinhold, 1967). Both views may be correct since the size and

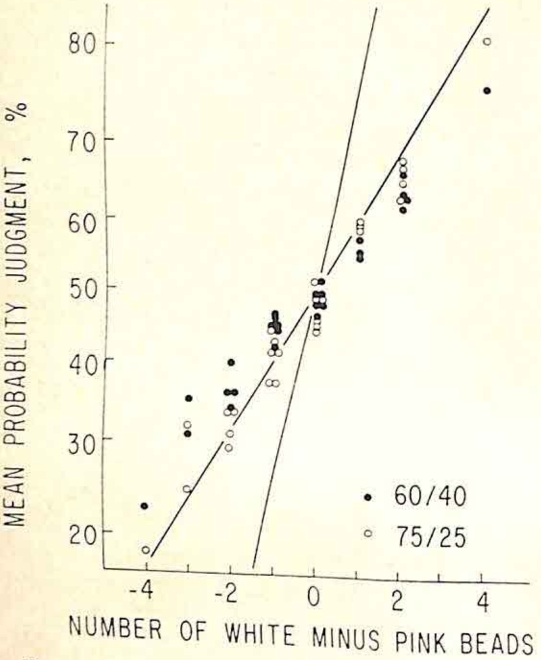


Fig. 2. Mean judgments for 60/40 and 75/25 conditions: additivity analysis for Bayesian model. (Horizontal axis is the difference in white and pink beads; vertical axis is the log-odds spacing of probabilities. Normative predictions appear as straight lines: the steeper line is the 75/25 prediction, and the shallower line is the 60/40 prediction.)

reliability of the effect depends on the particular serial positions analyzed.

*Additivity.*—The raw responses were analyzed both graphically and statistically for additivity. The graphical analyses in Fig. 1 employ the backward projection scheme proposed by Shanteau (1970). The closed circles in each tree diagram were produced by successive averaging of the R4 values; e.g., each R3′ point is the average of the two connected R4 points. Additivity means that all upward (or downward) sloping lines between adjacent serial positions should be parallel. Inspection of Fig. 1 reveals some evidence of nonadditivity between R3′ and R4, but the remainder of the plots appear quite orderly.

Statistically, the integration model predicts that all four three-way interactions in the $2^4$ factorial should be nonsignificant. Analysis of the probability responses revealed that none of these interactions were significant. There was one discrepant two-way interaction in the 75/25 condition between the third and fourth (3 × 4) serial positions, $F$ (1, 23) = 4.65. However, the two-way and four-way interactions are of no interest since symmetry alone (between pink and white beads) implies that they should be zero. Since that means that nonsignificant interactions cannot be taken as support for the model, it also means that such interactions are not particularly relevant (Friedman, Carterette, & Anderson, 1968).

The Bayesian additivity prediction can also be analyzed graphically and statistically. The graphical analysis shown in Fig. 2 has been used in most normative work. The mean R4 and R9 values are plotted as a function of log odds on the vertical axis and white-minus-pink beads on the horizontal axis. Additivity implies that the data should plot as a straight line, and both the 60/40 (closed circle) and 75/25 (open circle) points appear to be roughly linear. For comparison, the normative predicted lines for the 60/40 and 75/25 conditions are also shown, and the conservatism effect is clearly evident, especially for the steeper 75/25 line.

To examine the Bayesian additivity prediction statistically, the data were first transformed into log odds and then analyzed as above. Two of the three-way interactions (Positions 1 × 2 × 4 and 2 × 3 × 4) were significant in the 60/40 condition, and one three-way interaction (1 × 2 × 4) was significant in the 75/25 condition. Overall, the log-odds data were noticeably less additive than the raw responses.

*Serial position curves.*—Since the integration model was supported, it can be used to estimate the serial weights. These estimates (from Equation 3 below) have been plotted in Fig. 3 as serial position curves. The R4:75/25 curve, for example, gives the influence of the first four beads on the fourth 75/25 response; thus, the weight of the third bead was just under 16%.

A recency effect is visible in both R4 curves, showing that later events in a sequence carry more impact. A similar trend can be seen in the R9 curves, although the 75/25 values are somewhat irregular. Some caution should be exercised, however, in interpreting the

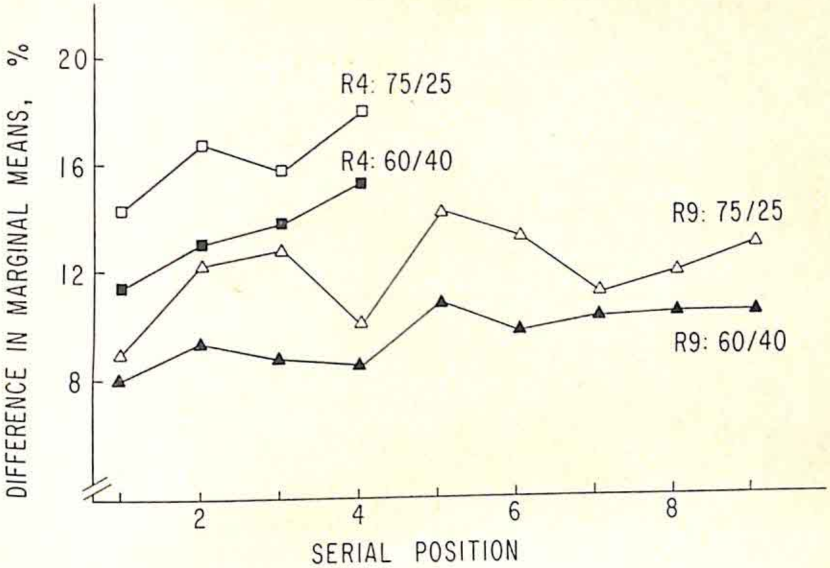


FIG. 3. Serial position curves of R4 and R9 judgments for 60/40 and 75/25 conditions.

R9:75/25 curve since an additivity analysis of the confounded $2^9$ design (Hahn & Shapiro, 1966) revealed a significant discrepancy, $F\ (2,\ 276) = 10.73$. Therefore, this serial curve may not be entirely valid. In contrast, there was no significant discrepancy for the R9:60/40 judgments, and so these values in Fig. 3 are valid weight estimates.

To check the reliability of the recency effects, the R4 weights were computed for each *S*. This permitted analysis of the weights across serial position, and significant order effects were found in both conditions. Essentially the same pattern of serial effects was found for the weights derived from the log-odds data.

*Weight of initial opinion.*—The initial weight values for the R4:60/40 and R4:75/25 judgments were .466 and .353, respectively. Thus, even after four samples, *Ss* were still attaching considerable importance to their initial opinion. These $w_0$ values also provide an index of the previously noted diagnosticity effect.

The remainder of this section summarizes the parameter estimation procedures. Each serial position weight was estimated by taking the difference, $D_i$, in marginal means for the corresponding factor of the design. In terms of Equation 2,

$$D_i = w_i(s_W - s_P). \qquad [3]$$

These differences are proportional to the weights and have been plotted in Fig. 3 (for further details, see Weiss & Anderson, 1969).

The estimation of the initial weight required some new procedures. Since the sum of the weights equals one,

$$w_0 = 1 - \sum_{i=1}^{n} w_i.$$

Plugging this into Equation 3 yields

$$D_0 = w_0(s_W - s_P) = (s_W - s_P) - \sum_{i=1}^{n} D_i. \qquad [4]$$

When $(s_W - s_P)$ is known, this can be used to estimate $w_0$.

The initial scale value can be estimated by using the grand mean, $m$. This can be expressed as

$$m = w_0 s_0 + \sum_{i=1}^{n} \frac{w_i}{2}(s_W + s_P).$$

Rewriting and substituting in Equation 3 leads to

$$w_0 s_0 = m - \frac{s_W + s_P}{2(s_W - s_P)} \sum_{i=1}^{n} D_i. \qquad [5]$$

The $s_0$ value can be derived directly by using Equation 4 to estimate $w_0$.

These estimation equations were used in the present study by letting $s_W = 1$ and $s_P = 0$.

## DISCUSSION

Of the three comparisons between the models, the data favored the integration model on additivity and serial effects, and agreed with both models on diagnosticity. However, the primary concern is not as much in the exact fit of the models as it is in deciding which provides a better basis for understanding inference judgments.

The greatest difference between the models was observed for the serial position effects. The recency effect replicates earlier studies (Pitz & Reinhold, 1968; Shanteau, 1970), and taken together these findings provide strong evidence of order effects in inference judgments. Of course, such effects are commonly observed in other information-processing tasks (Slovic & Lichtenstein, 1971). The integration model expects such effects and handles them with an assignment of weight parameters to the serial positions. The normative model, on the other hand, predicts equal weighting and so is unable to handle any order effect at all.

In terms of additivity, the integration model was definitely superior to the Bayesian model. The Bayesian model showed three out of eight significant interactions, while the integration model showed none. This result replicates Shanteau's (1970) finding that raw probability responses are additive. Together, these studies suggest that any nonlinear transformation, such as log odds, is inappropriate for probability responses. Of course, other response scales may lead to different results.

A strong diagnosticity effect was predicted by the Bayesian model, and a weak effect was found. However, within the constraints of additivity, the Bayesian predictions can be modified by the accuracy ratio to account for diagnosticity. The integration model, on the other hand, handled the observed effect through the initial weight parameter. While both models can handle such effects, the integration model does so directly through a descriptive analysis, while the Bayesian model does so through modification of a normative prediction. Thus, the integration model provides a more direct analysis of the variables which influence diagnosticity.

In each of these comparisons, the Bayesian model made exact, a priori predictions. And in each case, these predictions were incorrect. The integration model, on the other hand, did not "predict" any of the effects. However, it does allow for and is able to handle the results. The key point is that it can *measure* these effects, and these measurements are *validated* within the functional measurement framework (Anderson, 1971). Of course, these measurements can then be used by the model to predict other aspects of the behavior, such as the response on any trial to a particular sequence.

An additional advantage of the integration model is its ability to handle probability learning (Friedman et al., 1968) and estimation judgments (Shanteau, 1970) as well as inference judgments. The integration model thus provides a unified theoretical treatment of sequential decision making which is beginning to have considerable experimental support.

## REFERENCES

ADDELMAN, S., & KEMPTHORNE, O. *Orthogonal main-effect plans.* (Tech. Rep. No. ARL 79) Wright-Patterson Air Force Base, Ohio: United States Air Force, 1961.

ANDERSON, N. H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206.

BEACH, L. R., WISE, J. A., & BARCLAY, S. Sample proportion and subjective probability revisions. *Organizational Behavior and Human Performance*, 1970, **5**, 183–190.

EDWARDS, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment.* New York: Wiley, 1968.

FRIEDMAN, M. P., CARTERETTE, E. C., & ANDERSON, N. H. Long-term probability learning with a random schedule of reinforcement. *Journal of Experimental Psychology*, 1968, **78**, 442–455.

HAHN, G. J., & SHAPIRO, S. S. *A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments.* (TIS Rep. No. 66-C-165) Schenectady, N. Y.: General Electric Company, 1966.

PITZ, G. F., DOWNING, L., & REINHOLD, H. Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology*, 1967, **21**, 381–393.

PITZ, G. F., & REINHOLD, H. Payoff effects in sequential decision-making. *Journal of Experimental Psychology*, 1968, **77**, 249–257.

SHANTEAU, J. C. An additive model for sequential decision making. *Journal of Experimental Psychology*, 1970, **85**, 181–191.

SLOVIC, P., & LICHTENSTEIN, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, **6** 649–744.

WEISS, D. J., & ANDERSON, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52–63.

(Received July 12, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 69–71

# SCANNING, CHUNKING, AND THE FAMILIARITY EFFECT IN TACHISTOSCOPIC RECOGNITION[1]

D. J. K. MEWHORT[2]

*Queen's University*

The experiment measured memory load imposed by familiar and by random letter sequences both immediately after tachistoscopic presentation and after a 1-sec. period of organization. While *S*s could always report more letters from the familiar strings, the load imposed by familiar strings was greater than by random strings immediately after presentation. However, after a period of organization the latter relation reversed. The data were discussed as evidence for two familiarity effects, one associated with the transfer from iconic to short-term memory and one associated with more general organizational processes within short-term memory.

When a row of letters is presented tachistoscopically, the number of letters which can be reported correctly depends on the structure of the letter sequence. For example, more letters can be reported correctly from a nonsense row constructed to include frequent letter patterns than from a randomly derived row (Miller, Bruner, & Postman, 1954). Although a variety of authors has suggested a response bias or guessing interpretation of the familiarity effect, a recent series of experiments has shown that such an account is not adequate (Bryden, 1970; Mewhort, 1967, 1970). Instead, it appears that a large part of the benefit available for familiar patterns reflects ease of encoding into short-term memory. For example, Mewhort, Merikle, and Bryden (1969) have argued that the familiarity effect reflects processes involved in the transfer from iconic to short-term memory. Specifically, it was argued that the transfer process involves a left-to-right operation (scanning) which proceeds more rapidly the more familiar the letter pattern. Presumably, the more rapid the scanning, the more complete the input to short-term memory.

The claim that letters of a familiar string are easier to report because they are transferred more rapidly from iconic to short-term memory, however, appears to be at odds with data reported by Baddeley (1964). Baddeley presented random and familiar letter strings at a duration long enough for *S*s to read each letter aloud. Thus, his conditions preclude any useful scanning effects—the presentation time was too long for such effects to be relevant. His results showed, however, a large difference in recall favoring the familiar sequences. Baddeley's interpretation focused on the concept of chunking (Miller, 1956) and argued that familiar sequences are sustained more easily because they lend themselves more rapidly to organizational processes in memory.

Baddeley's work illustrates the incompleteness of the scanning model suggested by Mewhort et al. (1969). The two suggestions, however, are not incompatible. The present experiment was designed to explore the possibility that the familiarity effect reflects two processes, one associated with the transfer from iconic memory and another associated with more general organizational processes within short-term memory. The suggestion is that the two sources of the familiarity effect occur independently and successively. The strategy used in the experiment involved measuring memory load imposed by familiar and by random sequences both immediately after scanning and after a period of organization. Briefly, the argument is that because of more rapid transfer from iconic memory

---

[1] The research reported here was supported by the National Research Council of Canada Grant APA-318. The author thanks Pat Crawford for her assistance throughout the investigation.

[2] Requests for reprints should be sent to D. J. K. Mewhort, Department of Psychology, Humphrey Hall, Queen's University, Kingston, Ontario, Canada.

to short-term memory, familiar letter strings rather than randomly derived onces ought to impose a greater load on short-term memory immediately after scanning. Once in short-term memory, however, the material can be organized. Presumably, organization in memory is a skill requiring both time and central processing capacity (cf. Moray, 1967). Following Baddeley, it seems reasonable to assume that familiar sequences are organized more easily than random ones. Thus, once the materials have undergone organization within short-term memory, the relative load on memory for the two types of material ought to reverse.

## METHOD

*Subjects.*—The 16 *S*s participating in the experiment were undergraduates enrolled in introductory psychology courses at Queen's University. Participation in the experiment fulfilled part of the course requirement.

*Materials.*—Eight pseudowords of zero-order approximation and eight of fourth-order approximation were taken from the lists provided by Miller et al. (1954). Sixteen eight-item digit series were constructed by selecting digits randomly, subject to the restriction that no digit was repeated within a series.

The materials were prepared for presentation in a Gerbrands three-field tachistoscope by printing each sequence in uppercase letters on a white card in Letraset Instant Lettering (Style No. 287). The visual angle subtended by a row of items was approximately .5° × 4.0°. The letters were positioned approximately .5° above the horizontal of the visual field and the digits were positioned an equal distance below it. A fixation stimulus was prepared by placing two rows of ampersands in positions corresponding to the letters and digits. All materials were displayed with a luminance of approximately 4 ftl.

*Procedure.*—To begin each trial, a fixation field consisting of two rows of ampersands was exposed. When ready, *S* triggered the stimulus sequence. The exposure sequence followed (in order): a pseudoword of either zero-order or fourth-order approximation to English, a digit series, and finally a return to the fixation field. For half of the trials within each order of approximation, the time interval between the letters and digits was 0 sec. For the remaining trials, the interval was 1 sec. During the 1-sec. interval, the tachistoscope was dark. The exposure time for both the letters and the digits was 100 msec., and the fixation field remained on after exposure of the letters and digits for 2 sec.

The *S*s were instructed that they would be shown nonsense letter strings and random digit strings. In addition, they were told that some of the letter strings would look more like English than the others. In elaboration, *S*s were told about the nature of fourth-order pseudowords and were cautioned that none of the "familiar" material had been prepared by mutilating a real word. The *S*'s task was to report orally as many of the items as he could see. Order of report to be used required digits to be followed by the letters. The instructions emphasized that both the letters and the digits were "important," i.e., that a strategy of reporting only one or the other was not acceptable. Notice that although the letters were presented before the digits, recall of the digits preceeded recall of the letters. Thus, the load on memory imposed by the letters was present during presentation and recall of the digits, and changes in the load produced by varying the interstimulus interval could be monitored indirectly in terms of digit recall (cf. Savin & Perchonock, 1965).

Each *S* received 16 trials, i.e., four replications for the factorial combination of two variables, interstimulus interval, and order of approximation. Across *S*s, each pseudoword was used equally often for both interstimulus intervals. Pairing of pseudowords to the digit sequences was determined randomly for each *S*, and the order in which the conditions were presented was determined randomly for each *S*.

Before beginning the experiment proper, each *S* received a number of practice trials to familiarize him with the tachistoscopic task. Practice trials involved recognition of first-order approximations to English and started with a long exposure duration, e.g., 500 msec. Practice was continued with decreasing exposure until *S* could report at least four letters correctly at any exposure of 50 msec.

## RESULTS AND DISCUSSION

The number of letters and digits reported correctly was tallied for each trial. The data were collapsed across like cases and were entered into analysis of variance. The data are summarized in Table 1.

*Letter accuracy.*—As is shown on the right side of the table, more letters were

TABLE 1

PERCENTAGE OF ITEMS CORRECT

| Interstimulus interval | Material | | | |
|---|---|---|---|---|
| | Digits | | Letters | |
| | 0[a] | 4[a] | 0[a] | 4[a] |
| 0 sec. | 39.6 | 37.1 | 21.7 | 33.2 |
| 1 sec. | 42.6 | 46.5 | 27.9 | 46.9 |

[a] Approximation to English of letters.

reported correctly from the fourth-order material than from the zero-order material, $F\ (1, 15) = 41.257$, $p < .01$. In addition, recall was greater for both orders or approximation at the longer interstimulus interval, $F\ (1, 15) = 16.709$, $p < .01$, but there was no interaction between the variables, $F\ (1, 15) = 1.453$. In short, the data for letters showed a healthy familiarity effect throughout.

*Digit accuracy.*—As is shown on the left side of the table, the pattern of results for digits is quite different from that for the letters. Specifically, while more digits were reported correctly at the long interstimulus interval, $F\ (1, 15) = 7.035$, $p < .01$, there was a significant interaction of interstimulus interval with order of approximation, $F\ (1, 15) = 7.680$, $p < .01$. Thus, at the zero interstimulus interval, digit recall was lower when paired with a fourth-order approximation to English than with a zero-order approximation to English. At the 1-sec. interstimulus interval, however, the relative performance reversed.

The argument suggested earlier is that the load on short-term memory immediately after scanning ought to be greater for fourth-order material than for zero-order material, but that once the material has been organized, the relative load ought to reverse. In the present experiment, the load on memory imposed by letter sequences was monitored both immediately and after a 1-sec. period of organization in terms of the number of digits *S*s could identify. The pattern of results is consistent with the argument derived from memory load considerations. Thus, although *S*s could report more letters from a fourth-order string than a zero-order string at both interstimulus intervals, at the zero interstimulus interval recall of digits paired with fourth-order material was lower than of those paired with zero-order material. At the longer interval, however, the latter relation reversed. Thus, the experiment provides evidence that scanning is not identical to chunking but that the two processes are complementary.

## REFERENCES

BADDELEY, A. D. Immediate memory and the "perception" of letter sequences. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 364–367.

BRYDEN, M. P. Left–right differences in tachistoscopic recognition as a function of familiarity and pattern orientation. *Journal of Experimental Psychology*, 1970, **84**, 120–122.

MEWHORT, D. J. K. Familiarity of letter sequences, response uncertainty, and the tachistoscopic recognition experiment. *Canadian Journal of Psychology*, 1967, **21**, 309–321.

MEWHORT, D. J. K. Guessing and the order of approximation effect. *American Journal of Psychology*, 1970, **83**, 439–442.

MEWHORT, D. J. K., MERIKLE, P. M., & BRYDEN, M. P. On the transfer from iconic to short-term memory. *Journal of Experimental Psychology*, 1969, **81**, 89–94.

MILLER, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

MILLER, G. A., BRUNNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology*, 1954, **50**, 129–139.

MORAY, N. Where is capacity limited? A survey and a model. *Acta Psychologica*, 1967, **27**, 84–92.

SAVIN, H. B., & PERCHONOCK, E. Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 348–353.

(Received July 12, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 72–82

# VISUAL PROCESSING CAPACITY AND ATTENTIONAL CONTROL[1]

RICHARD M. SHIFFRIN[2] AND GERALD T. GARDNER

*Rockefeller University*

Three experiments tested whether visual processing operates under attentional control, and with temporal-spatial capacity limitations. The *S*s identified which of two key letters was present in briefly presented four-letter displays. The simultaneous condition presented the letters concurrently for *t* msec., preceded and followed by masking fields. The sequential condition presented the letters successively, each preceded and followed by a masking field, each for *t* msec. In the sequential condition, *S*s were given the onset order of the four letters. Models postulating attentional control and limited capacity would predict an advantage for the sequential condition since in this case processing capacity need not be simultaneously shared among four letters. The results demonstrated simultaneous and sequential conditions to be equal. It was concluded that the initial stages of visual processing, up to at least the level of letter recognition, take place without capacity limitation and without attentional control.

This paper presents three experiments designed to test two basic questions about visual information processing. First is the question of *attention*: can the amount and rate of information processed from a briefly presented image at a given location in the visual field be varied by *S* voluntarily (excluding muscluar control such as eye movements, focusing effects, and postural adjustments)? Second is the question of *capacity*: is the information processed from a given location of a briefly presented display affected by the simultaneous processing of information from other locations in the display? The most well-known quantitative models of visual processing, those of Estes and his co-workers (Estes & Taylor, 1964, 1966; Estes & Wessel, 1966) and that of Rumelhart (1970), imply a yes answer to both questions. On the other hand, recent empirical and theoretical work by Eriksen and Spencer (1969) and Gardner (1970, 1972) argue strongly that the answer to both is no.

We turn to a summary of this work (see Gardner, 1972, for a more complete review) and begin by defining the two most relevant tachistoscopic paradigms. When a display of symbols, usually letters, is presented tachistoscopically and *S* asked to try to report all the symbols presented, the procedure is termed "whole report" (Sperling, 1960). When *S* is instead asked to report merely the presence or absence of a known letter or to state which of two known letters is present, then the procedure is called "detection" (Estes & Taylor, 1964, 1966).

We consider first the question of capacity limitation. All researchers agree that capacity is limited to at least a certain degree. The agreement results from the extensive documentation of lateral masking: when information is presented in closely adjacent spatial locations, interference occurs and the processing from both locations is hindered (for example, see Collins, 1969; Woodworth & Schlosberg, 1954, pp. 104–105). However, when the locations used are sufficiently separated spatially, the capacity question is a very real one. The evidence in favor of limited capacity is centered in the *display size effect*: *S*'s probability of reporting any given letter in the whole-report procedure and *S*'s probability of correct judgment in the detection procedure both fall as the number of letters in the display increases (Estes & Taylor, 1964, 1966; Sperling, 1960). These effects are easy to explain if the amount and rate of information processed from a given display location is smaller when the display contains a larger number of letters. An alternative explanation of the display size effect is available, however. Sperling (1960, 1963) suggested that the performance dropoff in the whole-report procedure was a result of well-known postperceptual short-term memory limitations; i.e., all the letters could be processed and entered into memory, but not all could be successfully recalled from memory thereafter (see also Atkinson & Shiffrin, 1968). The detection procedure was designed by Estes to bypass these memory limitations; and hence the short-term memory explanation would not hold for the display size effect found in this paradigm.[3] However, the display size effect can be predicted from the effect of confusions among letters on the decision process *S* must use. This idea was first proposed by Eriksen and Spencer (1969) for their yes–no detection experiments and extended by Gardner (1972) to detection paradigms in general. The effect of confusions, in this view, can be outlined as follows. The more letters there are in a display, the greater are the number of letters that will be confused with the letters to be detected. The greater the number of letters confusable with the letters to be detected, the more difficult it will be to decide which of the two key letters was in fact presented. Indeed Gardner (1971) has shown that performance does not decrease with display size if the distractor elements in the display are not confusable with the letters to be detected. Also, Estes (personal communication) has reported the results of an experiment with numerous conditions, two of which replicated the finding that a display size effect disappears when the distractor elements are not confusable with the target elements. This experiment used linear arrays of four, six, and eight elements.[4]

---

[1] The research in this paper was supported by United States Public Health Service Grants GM 1789 and MH 12707-05. The authors are indebted to David Allmeyer for running the experiments, helping in the analysis of the data, and providing numerous helpful insights.

[2] Requests for reprints should be sent to Richard M. Shiffrin, now at the Department of Psychology, Indiana University, Blomington, Indiana 47401.

The evidence in favor of controlled attention derives from experiments using a postexposure cue (Averbach & Coriell, 1961; Sperling, 1960). At some point following display offset, a cue is given to *S* telling him to report only the letters in a specified subset of display locations. Recall of the subset is superior to recall of the entire display, but the difference goes to zero as the cue delay approaches .5 sec. Attention models explain this result by assuming *S* shifts his attention entirely to the cued subset at the moment of cuing; the smaller the subset, the greater attention will each member receive (see Rumelhart, 1970). An alternative explanation would assume that postperceptual memory effects would be increasingly bypassed as the cued set became smaller. This explanation does not posit switching of attention. Although postexposure cuing has not been attempted with a detection procedure, both the attention model and the confusion-decision model mentioned above would predict an advantage for a smaller cued set; such an experiment would therefore be indecisive on this issue.

The Eriksen and Spencer (1969) experiments provide evidence for "unlimited"-capacity, nonattention models of visual processing. These experiments used a circular array of 10 positions and a detection technique. The 10 letters of the display were presented 1 at a time in the 10 locations, in a random sequence. The critical variable in these experiments was the time between successive letters (ISI). The *E*s hypothesized that when the ISI was very short, say 5 msec., any limitations on processing capacity would come into play and processing capacity would have to be shared among the 10 letters presented. On the other hand, when the ISI was long, say 3000 msec., each letter could be processed separately and no processing limitations should be encountered. Therefore, capacity limitations should lead to superior performance at the very long ISIs. To sum up their results, detection performance was identical for ISIs of 5 msec. and 3,000 msec. Although these results are quite striking, a number of methodological problems in the experiments prevent us from concluding without question that per-

---

[3] In many whole-report and detection experiments, both lateral masking and acuity factors were not sufficiently controlled. One might argue that the results taken to indicate capacity limitations in these tasks were due to these uncontrolled factors. However, Gardner (1970) controlled for both lateral masking and acuity and failed to observe differences from the previously reported patterns of results.

[4] We refer the reader to Gardner's (1971) paper for a complete discussion of the effects of confusability.

TABLE 1

Results of Experiment I

| Level of confusability and cond. | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $\bar{X}$ |
|---|---|---|---|---|---|---|---|
| a. Probability of correct letter detection | | | | | | | |
| Low | | | | | | | |
| Simultaneous | .86 | .84 | .72 | .89 | .87 | .70 | .818 |
| Sequential | .78 | .77 | .72 | .89 | .84 | .71 | .785 |
| Difference | .08 | .07 | .00 | .00 | .03 | −.01 | .033 |
| High | | | | | | | |
| Simultaneous | .61 | .66 | .65 | .79 | .76 | .59 | .683 |
| Sequential | .67 | .55 | .63 | .75 | .77 | .59 | .667 |
| Difference | −.06 | .11 | .02 | .04 | −.01 | .00 | .016 |
| $M_{\text{diff.}}$ | .01 | .09 | .01 | .02 | .01 | −.01 | .025 |
| b. Probability of correct location judgment | | | | | | | |
| Low | | | | | | | |
| Simultaneous | .82 | .90 | .70 | .96 | .85 | .82 | .833 |
| Sequential | .73 | .80 | .67 | .91 | .81 | .86 | .787 |
| Difference | .09 | .10 | .03 | .05 | .04 | −.04 | .046 |
| High | | | | | | | |
| Simultaneous | .40 | .48 | .36 | .61 | .58 | .32 | .469 |
| Sequential | .44 | .45 | .37 | .58 | .56 | .34 | .465 |
| Difference | −.04 | .03 | −.01 | .03 | .02 | −.02 | .004 |
| $M_{\text{diff.}}$ | .02 | .07 | .01 | .04 | .03 | −.03 | .025 |
| c. Joint probability of correct letter and location | | | | | | | |
| Low | | | | | | | |
| Simultaneous | .71 | .79 | .56 | .88 | .78 | .62 | .722 |
| Sequential | .64 | .67 | .55 | .85 | .74 | .66 | .680 |
| Difference | .07 | .12 | .01 | .03 | .04 | −.04 | .042 |
| High | | | | | | | |
| Simultaneous | .31 | .43 | .32 | .58 | .54 | .28 | .421 |
| Sequential | .35 | .41 | .33 | .54 | .53 | .27 | .420 |
| Difference | −.04 | .02 | −.01 | .04 | .01 | .01 | .001 |
| $M_{\text{diff.}}$ | .01 | .06 | .00 | .03 | .02 | −.02 | .022 |

Note—Number of observations are as follows: 1, 062 for $S_1$; 1,442 for $S_2$; 1,541 for $S_3$; 835 for $S_4$; 2,114 for $S_5$; and 800 for $S_6$. Mean stimulus durations in milliseconds are 47 for $S_1$; 35 for $S_2$; 30 for $S_3$; 43 for $S_4$; 50 for $S_5$; and 40 for $S_6$. Mean number of observations is 1,299 and mean stimulus duration is 41.1 msec. for all *S*s.

*sequential* condition, only one character was displayed at a time. The top diagram of Fig. 2 illustrates a typical stimulus exposure in the simultaneous condition: (*a*) the fixation point appeared continuously until *S* depressed a microswitch panel. When this occurred, (*b*) four masking patterns were displayed for 400 msec., one pattern at each of the four corner positions; each pattern consisted of a 4 × 6 dot matrix the same height and width as the stimulus letters. (*c*) Following these premasks, the four stimulus characters appeared for a preselected exposure duration, the duration determined individually for each *S* (see below). (*d*) Immediately after the offset of the stimulus array, four postmasks, identical to the premasks, were displayed continuously until *S* made his responses.

The second diagram of Fig. 2 (Exp. I) illustrates a typical stimulus exposure in the sequential condition: (*a*) the fixation point appeared continuously until *S*'s start signal, at which time (*b*) four premasks were displayed; 400 msec. after onset of the premasks, the upper-left-hand premask was replaced by a stimulus character that remained on for the same exposure duration used in the simultaneous condition. (*c*) After this duration had elapsed, the upper-left-hand character was replaced by a postmask, and simultaneously the upper-right-hand premask was replaced by the second stimulus character. The second character remained on for the preselected duration. This process continued in clockwise direction until all characters had been displayed. (*d*) Following the offset of the final character, the four postmasks remained on until *S* made his responses. Note that the exposure duration for any given character was identical in both simultaneous and sequential conditions.

*Procedure.*—Each *S* served in an initial practice session and three experimental sessions. All sessions involved the same basic procedure. On every trial, *S* centered his gaze on the fixation dot and initiated the stimulus exposure by pressing a microswitch panel. Following the exposure, *S* indicated whether the array contained a T or an F by depressing one of two panels and then indicated the location of the target by depressing one of four additional panels.

Trials were grouped in 16-trial blocks, successive blocks alternating between simultaneous and sequential conditions; a set of special marker stimuli was displayed for 500 msec. before each block to indicate the presentation condition for that block. Within each 16-trial block, half of the trials involved high confusability arrays and half involved low confusability arrays. For both types of array, T and F were used as target letters equally often and appeared at each possible spatial location with equal frequency. The level of confusability, choice of target letter, and location of target letter all varied randomly within a block, without prior warning to *S*. As a result of the experimental design, two successive 16-trial blocks represented a complete within-*S* replication of the experiment. The *S*s were permitted to work at their own pace and complete as many 32-trial replications as they could within each 2-hr. session; they were urged to stop and rest as often as necessary to avoid fatigue. The number of blocks completed per session varied from approximately 10 to 18.

During the practice session, *S*s were familiarized with the experimental procedure. They received trial-by-trial feedback on the first two 32-trial replications; feedback thereafter was only given informally at the end of the experimental sessions. Additional practice session trials were then used to determine a stimulus exposure duration for each *S* that yielded 80% accuracy in detecting the target letters in low confusability simultaneous arrays. The duration was determined by a modified up-and-down method and was altered occasionally during experimental sessions in order to maintain the 80% level despite changes in the degree of practice. These exposure duration changes, however, only occurred between complete, successive replications of the experimental design. The average exposure duration over all sessions is given for each *S* in Table 1.

The first 32-trial replication of each experimental session served as a warm-up period and the resulting data were not included in the summaries and analyses reported below.

### *Results*

The main features of the data are presented in Table 1a, 1b, and 1c, which give the results both for individual *S*s and for the group average. The table includes the data for the simultaneous and sequential conditions within each level of confusability and also averaged over confusability. The lowest horizontal row (Difference) gives the difference between simultaneous and sequential conditions in each case, a positive number indicating an advantage for the simultaneous condition. Table 1a gives the probability of correct letter detection. Table 1b gives the identical table except that the entries are the probability of a correct location judgment, and Table 1c gives the identical table except that the entries are the probability of a jointly correct judgment of location and letter.

The data are summarized by the lower-right-hand corner of each of Tables 1a, 1b, and 1c. This cell gives the overall difference between simultaneous and sequential conditions. In each case the simultaneous condition is marginally superior to the sequential condition. A repeated-measures analysis of variance (Lindquist, 1953) confirmed what is evident from this table. For Table 1a, both *S*s and level of confusability were significant, $F\ (5,\ 5) = 15.5$, $p < .005$, and $F\ (1,\ 5) = 48.5$, $p < .001$, respectively. Type of presentation (sequential or simultaneous) was far from significant, and none of the interactions were significant. This pattern was the same for Tables 1b and 1c.

An examination of the finer details of the data (not shown in the tables) revealed that the upper two display locations exhibited a higher level of correct responding on all measures than the lower two display locations, but this did not interact with mode of presentation or level of confusability. Also, most *S*s exhibited a marked response preference for either T or F, about half preferring each. Again, this effect did not interact with presentation mode or level of confusability.

In summary, level of confusability had an appreciable effect: performance was poorer when the level of confusability was high; mode of presentation showed no effect whatever (although simultaneous displays were slightly numerically superior to sequential displays).

### *Discussion*

Quite obviously, this experiment showed no advantage for the sequential method of presentation of a display. This lack of advantage occurred despite the following facts. For simultaneous displays, all four letters had to be processed during a single exposure duration, while in sequential displays each letter could be processed separately for this same exposure duration; thus the sequential procedure involved a total duration of presentation four times that of the simultaneous procedure. Furthermore, the sequential presentations were biased so as to give *S*s maximum opportunity to switch attention and process the letters one at a time: the letters always appeared in a fixed cyclic order known to *S*, and the sequential trials were blocked so that no element of surprise would be involved. It seems fair to conclude that *S*s cannot switch attention or control their processing rates from different spatial locations, at least if attention must be switched at rates as fast as those used in this study (about 40 msec. per letter). Also the data indicate that the processing of four letters simultaneously is as efficient as the processing of four letters presented one at a time. This appears to imply that processing of the letters is independent and without a capacity limitation, at least up to four letters. Note that the significant effect of levels of confusability was predicted by the confusability models discussed in the introduction.

These results both replicate and extend those of Eriksen and Spencer (1969). Even when the order of onset is known, sequential presentation is not superior to simultaneous presentation. Are the present results amenable to different conclusions than we have drawn? While the answer is undoubtedly yes, the alternative explanations we have been able to devise have been too complex to merit discussion here. We will return to the question of alternative models after presentation of the remaining experiments.

## EXPERIMENT II

Experiment II was designed to test the effect of the premasks. It could perhaps be argued that the premasks attracted attention to each corner of the display and made it difficult to switch attention thereafter. The next experiment tested this possibility by eliminating the premasks of Exp. I.

### *Method*

*Subjects.*—Five of the *S*s from Exp. I served in Exp. II (one male and four females). They were paid $4.00 for each 2-hr. session, and each took part in one 2-hr. session.

*Stimuli and equipment.*—Experiment II was identical to Exp. I except that no premasks appeared before the stimulus characters. The top diagram of Fig. 2 illustrates a typical stimulus exposure in the simultaneous condition, except that the premasks are not used: (*a*) The fixation point appeared continuously at the center of the CRT. (*b*) Four hundred milliseconds after *S*'s start signal, the four stimulus characters were displayed for a preselected exposure duration. (*c*) Immediately after offset of the stimulus array, four postmasks were displayed continuously until *S* made his response.

The third diagram of Fig. 2 (Exp. II) illustrates a typical stimulus exposure in the sequential condition: (*a*) The fixation point appeared continuously at the center of the CRT. (*b*) Four hundred milliseconds after *S*'s start signal, the upper-left-hand stimulus character was displayed for the same exposure duration used in the simultaneous condition. (*c*) After this duration had elapsed, the character was replaced by a post-mask and simultaneously the upper-right-hand stimulus character was displayed. The second character remained on for the preselected duration. This process continued in clockwise direction until all characters were displayed. Each postmask remained on once displayed, and the full set of four postmasks appeared until *S* made his responses.

*Procedure.*—All procedural details were identical to those of Exp. I.

### *Results*

The central results of Exp. II are given in Tables 2a, 2b, and 2c. Only group averages are shown because just one session of data was collected for each *S*. The same analysis of variance was carried out as in Exp. I. For Table 2a, level of confusability, $F\ (1,\ 4) = 9.2$, $p < .05$, and mode of presentation, $F\ (1,\ 4) = 12.8$, $p < .05$, were significant but *S*s and the interactions were nonsignificant. The same pattern was true for Tables 2b and 2c. A

TABLE 2

RESULTS OF EXPERIMENT II

| Level of confusability and cond. | Mean (5 *S*s) probability of detecting: | | | No. observations |
|---|---|---|---|---|
| | a. Correct letter | b. Correct position | c. Correct both | |
| Low | | | | |
| Simultaneous | .72 | .79 | .61 | 520 |
| Sequential | .68 | .70 | .54 | 520 |
| Difference | .04 | .09 | .07 | |
| High | | | | |
| Simultaneous | .67 | .55 | .43 | 520 |
| Sequential | .64 | .54 | .41 | 520 |
| Difference | .03 | .01 | .02 | |
| $M_{\text{diff.}}$ | .04 | .05 | .05 | |

Note.—Mean stimulus duration = 47 msec.

finer grained analysis of the data revealed results similar to those in Exp. I; since these did not interact with the main effects, they will not be discussed.

### *Discussion*

This experiment was carried out to check the possibility that the premasks "drew" attention to all display locations and prevented attention switching. The results showed a small but significant advantage for *simultaneous* displays when premasks were not used. In light of these results, it seems improbable that the results of Exp. I could have been due to an attention-fixating effect of the premasks.

Since a single 2-hr. session for each *S* demonstrated that the sequential displays were not superior, the experiment was terminated at this point. A clue to the small advantage for simultaneous displays was revealed in the comments of several *S*s. They noted the sequential procedure to be somewhat confusing in relation to their prior experience (Exp. I) since the number of filled locations increased from the beginning to the end of a display sequence.

## EXPERIMENT III

In each of the previous experiments, it could be argued that presentation rates for the sequential displays (approximately 40 msec/character) were too fast for attention switching to take place. The natural next experiment would slow the rate of presentation by placing a variable duration mask between the onset of the successive letters. Unfortunately, at slow speeds, this procedure would allow *S* to make eye movements and track the cyclic display. In order to bypass the eye-movement problem, we altered the procedure slightly for the sequential condition. Instead of presenting a sequence of four single letters, the two letters lying on one diagonal of the square display were presented first, and then the two letters lying on the other diagonal were presented after a delay. Since for each diagonal the natural fixation point was still the dot at the center of the display, we hoped that *S*s would not be tempted to use eye movement. This procedure still allowed comparison with simultaneous displays, with the advantage that a long delay could be introduced between the presentation of the two diagonals.

### *Method*

*Subjects.*—All *S*s from Exp. II served in Exp. III. In addition, three new *S*s (two male and one female) who had no prior tachistoscopic experience served in the experiment. Data from one of these new *S*s were later discarded after both his response pattern and his answers to postexperimental questioning revealed a tendency to deviate markedly from the fixation point.

*Stimuli and equipment.*—Simultaneous arrays in Exp. III were identical in format to those in Exp. I, as indicated in the top diagram in Fig. 2; sequential arrays were different in format. The bottom diagram of Fig. 2 (Exp. III) illustrates a typical stimulus exposure in the sequential condition. In the description that follows, the upper-left and lower-right corners of the array are together referred to as "Diagonal A," and the upper-right and lower-left corners as "Diagonal B." (*a*) The fixation point appeared continuously until *S*'s start signal, at which time (*b*) four premasks were displayed; 400 msec. after the onset of the premasks, the two premasks in one diagonal, Diagonal A in the illustration, were replaced by stimulus characters, which then remained on for the same exposure duration used in the simultaneous condition. (*c*) After this duration had elapsed, the two characters were replaced by postmasks and four masks remained on the screen for 500 msec. (*d*) After this 500-msec. period had elapsed, the premasks in the other diagonal (in this case, Diagonal B) were replaced by the remaining two stimulus characters. The second pair of characters remained on for the preselected duration, and (*e*) were then replaced by postmasks. The full set of four postmasks remained on until *S* made his responses.

*Procedures.*—A complete replication of the experimental design involved 64 trials—four 16-trial blocks, successive blocks alternating between simultaneous and sequential conditions. Marker stimuli appeared before each block to indicate the presentation condition for that block. The two simultaneous-condition blocks were identical to the simultaneous blocks in Exp. I. One of the sequential blocks involved arrays in which Diagonal A appeared first, and the other sequential block involved arrays in which Diagonal B appeared first. The marker stimuli appearing before the sequential blocks indicated the order of appearance of the diagonals.

Within each of the four 16-trial blocks, half of the trials involved high confusability arrays and half involved low confusability arrays. For both types of array, T and F were used as target letters equally often and appeared at each possible spatial location with equal frequency. The level of confusability, choice of target letter, and location of target letter all varied randomly within a block. The *S*s were permitted to work at their own pace and complete as many 64-trial replications as they could without fatigue during each 2-hr. session.

All other procedural details were the same as in Exp. I.

TABLE 3

Results of Experiment III

| Level of confusability and cond. | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $\bar{X}$ |
|---|---|---|---|---|---|---|---|---|
| a. Probability of correct letter detection | | | | | | | | |
| Low | | | | | | | | |
| Simultaneous | .87 | .72 | .91 | .91 | .77 | .66 | .81 | .800 |
| Sequential | .84 | .72 | .89 | .89 | .73 | .69 | .80 | .794 |
| Difference | .03 | .00 | .02 | .02 | .04 | −.03 | .01 | .006 |
| High | | | | | | | | |
| Simultaneous | .69 | .66 | .79 | .79 | .58 | .62 | .63 | .684 |
| Sequential | .66 | .65 | .86 | .83 | .58 | .62 | .67 | .703 |
| Difference | .03 | .01 | −.07 | −.04 | .00 | .00 | −.04 | −.019 |
| $M_{\text{diff.}}$ | .03 | .01 | −.02 | −.01 | .02 | −.02 | −.02 | −.006 |
| b. Probability of correct location judgment | | | | | | | | |
| Low | | | | | | | | |
| Simultaneous | .91 | .82 | .98 | .92 | .90 | .73 | .96 | .881 |
| Sequential | .88 | .80 | .93 | .86 | .95 | .68 | .97 | .852 |
| Difference | .03 | .02 | .05 | .06 | −.05 | .05 | −.01 | .029 |
| High | | | | | | | | |
| Simultaneous | .52 | .45 | .66 | .63 | .40 | .36 | .50 | .504 |
| Sequential | .48 | .45 | .76 | .73 | .33 | .48 | .53 | .550 |
| Difference | .04 | .00 | −.10 | −.10 | .07 | −.12 | −.03 | −.046 |
| $M_{\text{diff.}}$ | .04 | .00 | −.02 | −.02 | .01 | −.04 | −.02 | −.009 |
| c. Joint probability of correct letter and location | | | | | | | | |
| Low | | | | | | | | |
| Simultaneous | .83 | .63 | .89 | .86 | .71 | .52 | .78 | .737 |
| Sequential | .77 | .63 | .86 | .80 | .71 | .53 | .78 | .715 |
| Difference | .06 | .00 | .03 | .06 | .00 | −.01 | .00 | .022 |
| High | | | | | | | | |
| Simultaneous | .49 | .38 | .62 | .58 | .33 | .29 | .43 | .444 |
| Sequential | .42 | .39 | .74 | .68 | .30 | .38 | .47 | .491 |
| Difference | .07 | −.01 | −.12 | −.10 | .03 | −.09 | −.04 | −.047 |
| $M_{\text{diff.}}$ | .06 | .00 | −.04 | −.02 | .01 | −.05 | −.02 | −.012 |

Note.—Number of observations are as follows: 1,412 for $S_1$; 2,176 for $S_2$; 1,280 for $S_3$; 2,560 for $S_4$; 960 for $S_5$; 2,176 for $S_6$; 2,240 for $S_7$. Mean stimulus durations in milliseconds are 29 for $S_1$; 36 for $S_2$; 35 for $S_3$; 45 for $S_4$; 36 for $S_6$; and 40 for $S_7$. Mean number of observations is 1,829, and mean stimulus duration is 38.7 msec., for all *S*s.

*Results*

The main results are given in Table 3a, 3b, 3c. This table corresponds in all respects to Table 1a, 1b, and 1c. The sequential condition refers to the successive diagonal procedure. Since the order in which the two diagonals were presented had no observable effect, the two orders have been lumped together in the "sequential" results in the table. The same analysis of variance used for Exp. I and II revealed only level of confusability, $F\ (1, 6) = 44.2$, and *S*s, $F\ (5, 5) = 77.6$, both $p\text{s} < .001$, to be significant. As the table indicates clearly, the simultaneous and sequential conditions did not differ noticeably ($F < 1$). A fine-grained analysis of the data again revealed patterns like those found in Exp. I (superior performance on the two upper display locations, etc.), but since these again did not interact with the main variables, they will not be discussed.[5]

*Discussion*

These results show much the same pattern as those from Exp. I and II. Sequential and simultaneous display procedures did not give rise to appreciably different performance. In this experiment, *S*s had full knowledge of the order of the diagonals and had in addition a full ½ sec. between diagonals to "switch" attention. Nevertheless, performance was the same whether two or four letters must be processed simultaneously. We must conclude that no internal attention switching occurs during perceptual processing from the visual field.

---

[5] A complete fine-grained breakdown of the data for all three experiments is available from the authors upon request.

Are there any alternatives to this conclusion? In all three experiments the small advantage of simultaneous over sequential displays was less when the level of confusability was higher. While this interaction was not statistically significant in the experiments taken separately, the fact that it occurred in all three probably demonstrates it to be a real phenomenon. One type of attention model could explain this interaction by supposing that attention is switched entirely to the critical display location (in simultaneous displays) only when the critical location is identified. It is supposed that less information is required to identify the critical location in these experiments than to identify the letter in that location. (This supposition is supported by the data.) When all locations are confusable, the location identification time would be longer, attention switching would be less rapid, and the simultaneous displays would show poorer performance. Unfortunately, this explanation predicts equality of simultaneous and sequential displays only when confusability is zero and, with confusing letters present, predicts an advantage for the sequential procedure. Since the main effect in all experiments showed no such advantage, the attention-switching explanation of the interaction would appear specious. Although similar attention models could be proposed with additional countervailing processes just balancing out the predicted advantage of sequential displays, pursuit of such complex models seems fruitless in the absence of some direct evidence pointing toward perceptual attention switching. Note that the interaction in the data can be explained by the effects of short-term memory factors: although the experiment was designed to minimize short-term memory capacity and limitation problems, these factors may not have been entirely bypassed. Thus a small amount of forgetting may have occurred during the course of the sequential presentations, and short-term capacity may have been strained in the high-confusability simultaneous condition. These factors could have been of minimal magnitude and still have been responsible for the small interaction observed.

Although the results indicate that perceptual processing is not spatially limited, this conclusion must pertain to the range of one to four letters. Eriksen and Spencer (1969) obtained similar results with 10-letter arrays but generalization to arrays of this size is not quite possible due to the procedural difficulties in their experiments that were discussed earlier. It seems clear that capacity limitations will eventually appear as display size is increased due to the appearance of lateral masking and acuity factors. Furthermore, even in a detection task the short-term memory system will eventually become taxed if a large enough number of letters is presented simultaneously; that is, some of the processed information in short-term may be forgotten before a decision can be reached. On the other hand, some forgetting from both short- and long-term memory may occur when a large enough number of letters is presented sequentially. (As noted earlier, both these effects may have begun to evidence themselves even with the four letter displays of the present experiments.) These considerations have given us pause in planning experimental replications with larger displays. Note, however, that an expectation that short-term memory factors will appear with use of larger displays in no way detracts from the present theoretical conclusions. We have no present reason to believe in spatial capacity limitations for displays of any size (subject to lateral masking and acuity factors.)

None of the provisos in the above paragraph regarding capacity limitation, weak as they are, apply to the conclusions concerning attentional control. If *S* does not have the capability of switching attention among four letters presented sequentially (or conversely, splitting attention among four letters presented simultaneously), there is no reason whatever to expect such a capability to develop for larger displays. The conclusion that attentional control does not operate in visual-perceptual processing has some far-reaching consequences. For example, Neisser (1967) postulated "preattentive" mechanisms primarily because a process was needed to direct perceptual attention ("focal attention" in Neisser's terms). If perceptual attention is not operative then preattentive mechanisms appear to be unnecessary.

Our conclusions regarding capacity limitations and attentional control strictly apply to a level of processing complexity no higher than that of letter detection. Nevertheless, we feel that the entire processing system prior to short-term memory has these characteristics. It would be difficult to imagine a system, for example, in which attention could be switched so as to process information more efficiently for whole report, without the same shift in attention giving rise to enhanced detection performance. In other words, it is very unlikely that perceptual attentional control could evidence itself in whole-report procedures, given that it did not evidence itself in the present experiments. Of course, in higher level tasks like whole report "limitations of capacity" and "attentional control" might well appear due to the contamination of those tasks by short-term memory factors.

The model of visual processing we are led to by our results is rather different from the Rumelhart (1970) conception and the early models of the Estes group (Estes & Taylor, 1964, 1966). This model is similar to that proposed by Gardner (1972): information enters the visual field and is perceptually processed "without limitation" and without attentional control up to and including at least the level of letter recognition. The various features abstracted are entered at once into short-term memory where they are scanned and decisions are made. The limitations of visual processing, in this view, are primarily postperceptual (with the exception of lateral masking type phenomena). Such results as the display size effect are handled in this framework in terms of short-term memory limitations or in terms of problems inherent in the decision structure of the task. In conclusion, then, taking into account the present results, those of Eriksen and Spencer (1969) and those of Gardner, we feel that a strong case has been established for an initial visual processing system with no capacity limitations other than those related to lateral masking and no attentional control on the part of *S*.

## REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1968.

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, **40**, 309–328.

Collins, J. F. Independence of sensitivity on different foveal areas. Unpublished doctoral dissertation, University of Illinois, 1969.

Eriksen, C. W., & Spencer, T. Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology*, 1969, **79** (2, Pt. 2).

Estes, W. K., & Taylor, H. A. A detection method and probabilistic model for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964, **52**, 446–454.

Estes, W. K., & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. *Perception & Psychophysics*, 1966, **1**, 9–16.

Estes, W. K., & Wessel, D. L. Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. *Perception & Psychophysics*, 1966, **1**, 369–373.

Gardner, G. T. Spatial processing characteristics in the perception of brief visual arrays. (Tech. Rep. No. 23) Ann Arbor, Mich.: The University of Michigan Human Performance Center, 1970.

Gardner, G. T. Evidence for independent parallel channels in tachistoscopic perception. (Indiana Mathematical Psychology Rep. No. 72-4) Bloomington: Indiana University, 1972.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1953.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Rummelhart, D. E. A multicomponent theory of perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, **7**, 191–218.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74**(11, Whole No. 498).

Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, **5**, 19–31.

Sperling, G. The description of luminous calibration of CRO displays. In N. Moray (Ed.), *On-Line computing for behavioral science*. Brussels NATO Institute, 1971, in press.

Wolford, G. L., Wessel, D. L., & Estes, W. K. Further evidence concerning scanning and sampling assumptions of visual detection models. *Perception & Psychophysics*, 1968, **3**, 439–444.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. (2nd ed.) New York: Holt, 1954.

(Received July 12, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 83–89

# STIMULUS CODABILITY AND LONG-TERM RECOGNITION MEMORY FOR VISUAL FORM[1]

TERRY C. DANIEL[2] AND HENRY C. ELLIS

*University of New Mexico*

An experiment was conducted investigating the role of distinctive verbal labels and scaled stimulus codability on the temporal course of recognition memory for complex random shapes. Either immediately or after delays of 15 min. or 1 wk., *S*s were given a five-stimulus, forced-choice, shape-recognition test and a free recall test for retention of associated verbal labels. High stimulus codability, longer stimulus exposure, and verbal label training all resulted in superior recognition performance. Shape recognition was generally stable over the delays employed, while label recall showed a significant retention decrement. More important, the course of shape-recognition memory and label recall was found to be essentially independent. It was concluded that the effect of verbal labels on memory for form, even under conditions designed to maximize label use by *S*, is best attributed to processes occurring during the encoding or acquisition phase of the memory task.

It has been found that learning to associate verbal labels with shape stimuli enhances subsequent recognition memory for the shapes (e.g., Ellis & Muller, 1964; Ellis, 1968). Attempts to account for this finding may generally be grouped into two categories, one focusing upon the role of the label as a mediator of shape recognition at the time of the memory test (e.g., Price & Slive, 1970) and the other emphasizing the effects of the label on what is encoded during initial exposure of the shape stimuli (e.g., Ellis, 1968; Ellis & Daniel, 1971). It has been proposed by Price and Slive (1970) that a verbal association, recalled at the time of shape-recognition testing, serves to aid *S* in the retrieval of the associated shape. In contrast, Ellis and Ellis and Daniel present evidence indicating that the effect of verbal labels is better attributed to stimulus selection/encoding processes occurring during training than to mediational or retrieval processes operative at the time of memory testing. They assume that representative verbal labels direct *S*'s attention toward distinctive features of the stimuli during shape–label associative training, thus enhancing *S*'s encoding of the stimuli.

Although Ellis and Daniel (1971) showed conclusively that the availability of learned verbal labels is not a necessary requisite for shape recognition, there may still be circumstances in which *S*s do use verbal labels to mediate memory for shape stimuli at the time of recognition testing. A number of variables are known, for example, to affect the extent to which *S*s make use of *E*-assigned verbal labels in retention of shape stimuli (e.g., Clark, 1965; Ellis & Homan, 1968). In general, the effects of explicit verbal label training on shape recognition memory would seem to depend upon the a priori codability of the shapes. Presumably, as the difficulty of encoding a shape increases, *S*s would tend to rely more extensively on assigned verbal labels. If, as proposed by Price and Slive (1970), *S*s use the verbal label to mediate shape recognition, then, conditions which maximize *S*s' reliance upon the label should optimize any dependence between shape recognition and label availability at the time of the memory test.

The object of the present experiment was to investigate the interaction of stimulus-encoding difficulty and verbal label training on delayed recognition memory for random shapes. Two experimental variables, scaled-

---

[1] This research was an MA thesis of the senior author under the direction of the junior author. This research was supported by National Science Foundation Grant GB-7926.

[2] Requests for reprints should be sent to Terry C. Daniel, who is now at the Department of Psychology, University of Arizona, Tucson, Arizona 85721.

TABLE 1

RELEVANT VERBAL LABELS AND CODABILITY SCALE VALUES FOR STIMULI

| High codability | | | Low codability | | |
|---|---|---|---|---|---|
| Vanderplas and Garvin Shape No. | Label | Associative latency (sec.) | Vanderplas and Garvin Shape No. | Label | Associative latency (sec.) |
| 3 | Swan | 6.5 | 23 | Bird | 20.8 |
| 2 | Anteater | 7.6 | 27 | Dog | 21.1 |
| 1 | Fountain | 12.7 | 18 | Flag | 21.6 |
| 7 | Duck | 12.9 | 17 | Dancers | 21.7 |
| 26 | Woman | 14.3 | 19 | Rabbit | 22.0 |
| 5 | Rabbit | 14.8 | 11 | Mountain | 23.3 |
| 30 | Boat | 14.9 | 29 | Boat | 23.5 |
| 12 | Bird | 15.1 | 24 | Man | 27.8 |
| Group $\bar{X}$ | | 12.35 | Group $\bar{X}$ | | 22.75 |

stimulus codability and duration of stimulus exposure, were combined as a means of varying the difficulty with which *S*s could encode shape stimuli. Shapes were either high or low in codability and were presented for brief or long durations as a way of varying the ease of stimulus encoding.

If *S*s do rely upon verbal labels to mediate shape recognition, the level of recognition performance should be dependent upon the availability of the appropriate verbal labels at the time of memory testing. In addition, as the combined effects of reduced stimulus codability and shortened exposure duration lead *S* to rely more extensively upon *E*-assigned labels, the degree to which shape recognition depends upon label availability (recall) should be increased. Moreover, to the extent that either labels or shapes are forgotten over the retention intervals, there should be a close relationship between specific shapes and labels forgotten. Alternatively, if the effect of verbal labels is primarily on the initial encoding of the shape stimuli (Ellis, 1968; Ellis & Daniel, 1971), then increasing the difficulty of stimulus encoding should result in a greater advantage in training for *S*s provided with representative verbal labels. This advantage would be reflected in superior shape-recognition performance for label-trained *S*s. Introduction of longer retention intervals, while conceivably affecting the overall level of recognition performance, would not be expected to alter the relative position of label- and nonlabel-trained *S*s. Furthermore, whatever the retention interval, the availability of verbal labels at the time of testing would not necessarily be associated with superior recognition performance.

## METHOD

*Design and subjects.*—Three between-*S* variables, i.e., delay interval, pretraining type, and stimulus codability, were combined with one within-*S* variable, stimulus exposure time, into a 3 × 2 × 2 × (2) factorial design. The *S*s were given observation (O) or relevant verbal label (RL) pretraining with high or low codability stimuli and were then tested for recognition of the stimuli either immediately, 15 min., or 1 wk. after training. All of the *S*s saw one-half of the training stimuli for .5 sec. and the other half for 6.0 sec. Ten university student *S*s were randomly assigned to each of the 12 experimental groups. A total of 120 *S*s were paid $1.00 or received credit toward a course requirement for their participation.

*Stimuli and apparatus.*—Stimuli were 16 Vanderplas and Garvin (1959), 24-point random shapes scaled for codability (C). Stimuli were scaled by presenting each of the 30 Vanderplas and Garvin 24-point shapes to individual *S*s ($N = 38$) with instructions to provide a "name or verbal association" for the shape as rapidly as possible. Mean latency of the first associative response is the scale value for each shape, short latencies indicating high codable shapes.

The eight shapes scaled highest in C were designated high-C shapes and the remaining eight, which were scaled lowest in C, were designated low-C shapes. Relevant (modal) verbal labels were selected for each of the stimuli on the basis of previous scaling data (Ellis & Muller, 1964). Shapes, labels, and C-scale values (associative latencies) of the stimuli employed are presented in Table 1.

All stimulus materials were photographed and mounted on 35-mm. slides. The slides were projected onto a screen by two Kodak "Carousel" (700) slide projectors. Exposure times and interstimulus intervals were programmed on magnetic tape.

*Pretraining procedure.*—Stimuli, eight high- or eight low-C shapes, were presented one at a time in four different random orders. Four of the stimuli were exposed for .5 sec. on each presentation and four were exposed for 6.0 sec. The order of exposure durations, which was constant for all conditions, was randomly determined with the restriction that half of the exposures in each block of eight were .5 sec. The stimulus presentation order, and thus exposure duration, was randomized between *S*s with the restriction that each stimulus was exposed for the same duration on all four presentations to each *S*.

In the O condition, shapes were projected alone onto the screen and *S*s were instructed to observe and remember the eight shapes. For RL *S*s, the verbal label was shown beside each shape and *S* was instructed to associate the word with the shape so as to "help him remember the shape." For both .5- and 6.0-sec. stimuli, the printed label was presented alone on the screen for 1.0 sec. immediately after the shape–label exposure. Observation *S*s were presented with a blank slide during this 1.0-sec. interval.

All *S*s were told that they would be required to recognize the shapes in a test to be given immediately, 15 min. or 1 wk. after training. The recognition test was described in detail to all *S*s prior to the initial training.

*Retention interval.*—An equal number of *S*s were tested at each of the three retention intervals. For *S*s tested immediately after training, approximately 30 sec. elapsed before testing during which *S*s were refreshed on the recognition test instructions. The *S*s in the 15-min. delay conditions listened to a tape recording of a popular comedian's routines during the entire interval. The tape recording was chosen as a method of occupying *S* during the 15-min. delay interval. Informal observation of *S*s and verbal reports of pilot *S*s indicated that the recording was sufficiently engrossing to prevent overt rehearsal and at the same time would not appear to be likely to interact with any of the experimental treatments. The 1-wk. delay *S*s listened to 7.5 min. of the recording immediately after training and to the other 7.5 min. immediately before being tested. This was done in order to equate the stimulus context both immediately after training and immediately before testing for the delayed *S*s.

While independent groups of *S*s were tested after each delay interval, *S*s in the immediate- and 15-min.-delay conditions were also retested at subsequent retention intervals. Immediate test *S*s were retested after 15 min. and again after 1 wk., thus providing a complete within-*S* replication of the experiment. The 15-min.-delay *S*s were retested after 1 wk. The data for any *S* failing to complete the entire test sequence, up to and including the 1-wk.-delay test, was dropped from the experimental analysis.[3]

*Shape recognition task.*—The retention test employed was a five-stimulus, forced-choice, recognition test. Alternative shapes were constructed according to a procedure described by Ellis and Muller (1964). Each of the 24 points of a prototype shape (pretrained stimulus) was moved ½ in. (1.27 mm.) up, down, left, or right at random. The resulting points were then connected to yield a variation shape. One prototype and four ½-in. random variations of that prototype were projected side by side on the screen in front of *S*. All *S*s were instructed that each set of stimuli contained a prototype. The shapes in each set were numbered from left to right, and *S* responded by saying aloud the number below the shape that he thought he had seen before. Sets were exposed for 4.0 sec. with a 2.0-sec. interset interval, and the order of presentation was randomized for each *S*. The *S*s were instructed to respond to each set before the next set was exposed. Two blank exposures preceded the first test set in order to pace *S* on the task.

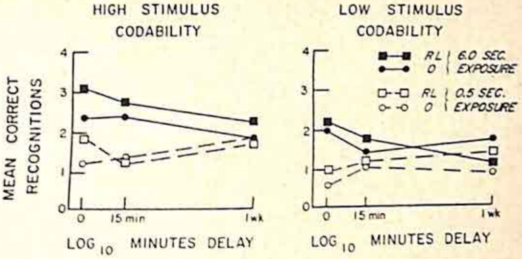


FIG. 1. Mean shape recognition for observation and relevant labeling conditions with .5-sec. and 6.0-sec. stimulus exposure as a function of postpredifferentiation delay.

*Label recall.*—Immediately following the shape-recognition task, all RL *S*s were required to write down as many of the pretraining labels as they could recall.

## RESULTS

*Recognition performance.*—Shape-recognition data from the independent groups experiment are presented in the two panels of Fig. 1. Mean correct recognitions are plotted against the amount of postpredifferentiation delay for high- and low-C stimuli. Separate retention curves are plotted for RL and O pretrained groups in which stimuli were exposed for either .5 sec. or 6.0 sec., respectively.

Although there appears to be some indication of a trend in one or two of the curves presented, overall there is no evidence of any systematic effect of postpredifferentiation delay on the course of recognition memory for random shapes. The average of the recognition performance over all high-C stimuli immediately after training is 53% (2.1 correct) and only drops to about 47% (1.9 correct) after a full week of intervening time. Immediate and 1-wk.

---

[3] A total of 17 *S*s were dropped from the main experiment for failure to return to the laboratory to complete the required test sequence. The range across cells was from two cases of no *S*s lost to one case of four *S*s lost. Where scores for the earlier tests were available, each *S* ($N = 12$) not returning was matched with a returning *S* selected at random from the same cell. A comparison of the performance on completed tests for these *S*s revealed no significant differences, $t$ (11) < 1.0.

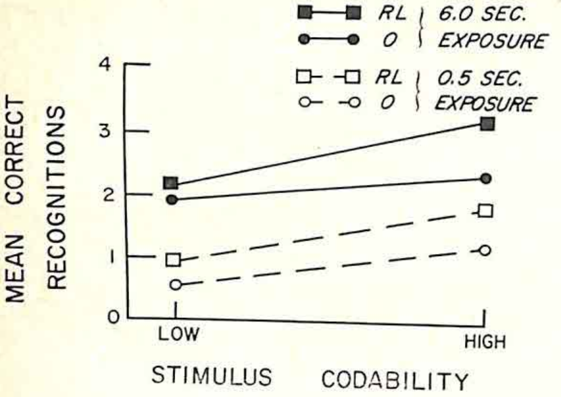


Fig. 2. Mean correct recognitions on the immediate (zero-delay) test as a function of stimulus codability, pretraining type, and stimulus exposure duration.

delay averages with low-C shapes, 34% (1.4 correct) and 32% (1.3 correct), again indicate that there is no decline in recognition accuracy over the delay intervals employed in the present experiment. This observation is supported by the results of the overall analysis of variance in which the main effect of postpredifferentiation delay did not approach significance, $F\ (2, 108) < 1.0$.

The analysis did, however, indicate a significant interaction of delay with the duration of stimulus exposure, $F(2, 108) = 8.69$, $p < .01$. Inspection of Fig. 1, however, reveals that a slight tendency for recognition performance to improve for stimuli exposed for .5 sec. in pretraining contributes as much to this effect as does the slight decrement observed for the stimuli exposed for 6.0 sec. Overall performance for .5-sec. stimuli rose from 28% immediately after training to 36% at 1 wk. of delay, while 6.0-sec. exposed stimuli showed a decline from 59% to 43% over the same interval. An analysis of the within-*S* replication data (immediate test group) supports the nonsignificance finding with respect to the main effect of delay, $F\ (2, 72) = 1.32$, $p > .05$, but does not indicate a Delay × Stimulus Exposure interaction, $F\ (2, 72) = 2.72$, $p > .05$.

Further comparison of the two panels in Fig. 1 reveals that recognition of stimuli scaled high on the codability measure, left panel, was generally superior to that for stimuli of low C. Both the between-*S* analysis and the within-*S* analysis support this observation, codability main effects being significant in both cases, $F\ (1, 108) = 25.87$ and $F\ (1, 36) = 20.83$, $ps < .01$, respectively. In neither analysis were interaction effects involving codability found to be significant at the .05 level.

In Fig. 2, data are plotted for the immediate recognition test alone in order to present the effects of stimulus codability and distinctive verbal label training as compared to observation practice. The general superiority of recognition accuracy for stimuli of high C is quite evident in Fig. 2, as is the lack of interaction effects with the other variables represented.

The influence of RL training is also evident in Fig. 2. Label-trained *S*s showed superior recognition performance for both high- and low-C stimuli with either of the two stimulus exposure durations. The overall analysis of the between-*S* experiment, including the delay variable, confirmed this observation yielding a significant main effect of pretraining type, $F\ (1, 108) = 4.36$, $p < .05$. There were no significant interactions involving type of pretraining in this analysis.

The pretraining main effect failed to reach significance in the within-*S* replication, $F\ (1, 36) = 1.12$, $p > .05$. This inconsistency seems to be the result of a tendency for these RL *S*s to show a slight retention decrement between the immediate and 1-wk. tests, dropping from 50% correct to 43% correct. This effect apparently combined with a slight tendency for *O*-trained *S*s to show an improvement in performance over the same intervals, going from 37% correct immediately after training to 42% correct after 1 wk. of delay. The result was a significant Pretraining × Delay interaction, $F\ (2, 72) = 3.27$, $p < .05$.

Stimulus exposure effects are evident in both analyses, 6.0-sec. exposures leading to consistently superior performance on the recognition task. This effect was highly reliable, $F\ (1, 108) = 66.21$ for the between-*S* analyses and, $F\ (1, 36) = 38.89$ in the within-*S* analysis, *p*s $< .01$. The only interaction involving exposure was that with delay, discussed above.

*Label recall.*—Mean number of labels correctly recalled are presented in Table 2 for all RL groups in the between-*S* experiment.

A general trend toward a decrement in label recall performance may be seen in the means of the delayed groups as compared with the immediate test groups. This trend proved significant in the overall analysis of variance for the main effect of delay, $F\ (2, 54) = 3.80, p < .05$. No other main effects or interactions approached significance in the label-recall analysis. It should be noted that at no point did the level of label recall fall below that of shape recognition for any of the experimental conditions.

While the overall level of label recall and shape recognition provides an index of shape–label association, a more direct measure would involve the relationship between memory for specific shape–label pairs. An *S*-by-*S* analysis of the correlation between the recognition of a shape stimulus and the subsequent recall of the label paired with that stimulus for all RL *S*s revealed very low, nonsignificant, positive correlations for both high and low codability forms, phi coefficients being .09 and .12, respectively.

In addition to the correlation analysis, a chi-square test of independence between recognition of a shape and recall of its appropriate label was also performed. This analysis revealed that while there was no evidence for a dependence relation in the case of high-C shapes, $\chi^2\ (1) = 2.06$, $p > .05$, there was some tendency for label recall and shape recognition to be related in the low-C condition, though again the statistic failed to reach significance, $\chi^2\ (1) = 3.71$, $p > .05$.

Separate chi-square tests for each Delay × Exposure × Codability combination again failed to support a shape-label dependence hypothesis. Only the .5-sec. exposure of low codability shapes at zero delay produced a significant relation, $\chi^2\ (1) = 7.94$, $p < .01$. Even in this case, the nature of the relationship was not in the appropriate direction. There appeared to be a disproportionate number of cases in the label recalled, shape-not-recognized cell. This was chiefly due to the fact that many more labels were recalled than shapes recognized in this condition.

TABLE 2

Mean Number of Labels Recalled for Distinctive Verbal Label Groups

| Codability | Delay interval | | | | | |
|---|---|---|---|---|---|---|
| | Zero | | 15 min. | | 1 wk. | |
| | Stimulus exposure | | Stimulus exposure | | Stimulus exposure | |
| | .5 | 6.0 | .5 | 6.0 | .5 | 6.0 |
| High | | | | | | |
| $\bar{X}$ | 3.0 | 3.2 | 2.4 | 3.5 | 2.7 | 2.2 |
| S | .82 | 1.03 | 1.08 | .22 | .30 | 1.23 |
| Low | | | | | | |
| $\bar{X}$ | 2.8 | 2.9 | 3.0 | 3.0 | 2.4 | 2.6 |
| S | 1.03 | .57 | 1.05 | .66 | 1.58 | .84 |

*Supplementary analyses.*—An analysis of the intra-*S* consistency in shape recognition was conducted for the *S*s in the within-*S* replication. These *S*s were tested once at each of the three delay intervals as described earlier in the method section. A conditional probability measure, the probability of *S*'s correctly recognizing a shape on both the 15-min. and the 1-wk. test given that he had recognized that shape correctly on the immediate test, revealed a high degree of consistency in shape recognition performance. Conditional probability was .66. The probability expected by chance in this case is .04.

## Discussion

One important feature of the results of this experiment was the lack of any overall effects of postpredifferentiation delay on recognition memory for random shapes. Finer analysis of correct recognitions for each stimulus by individual *S*s who were tested after each of the delay intervals also revealed a high degree of consistency in performance from test to test. These findings generally confirm those reported earlier (Clark, 1965; Ellis & Daniel, 1971) by showing the stability of recognition memory for random shapes over time.

In the present experiment, high stimulus codability, distinctive verbal label pretraining, and increased stimulus exposure independently produced superior levels of recognition performance. The stability of recognition memory over the delay intervals studied suggests that the effects of these variables probably should not be considered to have affected memory for the forms but rather to have affected the level of original learning or acquisition.

These results are quite consistent with an account of the verbal label effect offered by Ellis (1968). Ellis ascribed two principal functions to the distinctive verbal label in the shape predifferentiation situation. The first was a confirmation function, the label serving to inform *S* when he has minimally discriminated among the training forms. In addition, the label, when relevant, also may serve to direct *S*'s attention to specific characteristics of the stimulus, thus enabling increased learning of the distinctive features of the form.

The fact that associating relevant verbal labels with the form stimuli enhanced recognition performance would tend to support Ellis' (1968) account. While Ellis did not specifically treat stimulus exposure duration as a factor, the effects of this variable would also seem consistent with his conception. Presumably, the longer exposure would allow more time for learning distinctive characteristics of the forms. Codability effects would also fit into this account of verbal label effects. To the extent that high-C forms allow for more rapid association of the label (confirmation function), more time would be available for learning about additional distinctive characteristics of the shape (directing function).

On the surface, the fact that label training did not produce differential effects as a function of either codability or exposure would seem somewhat inconsistent with the Ellis (1968) formulation. In the present experiment, it was initially hypothesized that as the difficulty of encoding the form stimuli was increased the influence of the verbal label would be more pronounced. This hypothesis was not supported by the data. A possible explanation of this finding is suggested by Ellis' account of the differential effects of relevant labels on recognition performance as a function of the association value (AV) of the shape stimuli. Ellis reported that the effects of label relevance depended upon the AV level of the forms, relevance being relatively unimportant for low-AV forms. Ellis suggested that this may have been due to the fact that the relevant labels were less *representative* (i.e., would be given as associates less often) of the low-AV shapes than were the labels assigned to the high-AV forms. As a result, the directing function of the relevant label was greatly reduced when low-AV forms were used.

The similarity of the codability measure used in the present experiment to the AV measure used in the Ellis (1968) experiment would suggest a parallel analysis of the effects of label training and form codability. First, it was observed that RL training did result in generally superior recognition performance relative to O training in the low-C conditions. The difference between O and RL groups at low C was, however, no greater than that found to obtain under conditions of high C. Indeed, if any difference existed at all it appeared to have favored the high-C groups. This result would follow if it is assumed that the relevant labels for low-C forms were not sufficiently representative of those forms to enable full implementation of the directing function of the label.

In general, the effects of form codability, exposure duration, and verbal label training coupled with the observed stability of recognition performance over the delay intervals employed would tend to support accounts which emphasize the role of the verbal label in the initial learning or encoding of the form stimuli. It might be argued that these variables affected the amount or the nature of the information encoded during training, but, that information, once encoded, was equally available to *S* throughout the delay intervals.

The observed lack of association between label recall and shape-recognition performance would further suggest that availability of a verbal code at the time of testing is neither necessary nor sufficient for correct recognition of the shape stimulus. In this respect, these findings confirm those reported by Ellis and Daniel (1971) and extend those results by showing that this lack of association is maintained under conditions designed to maximize the importance of *E*-assigned labels. Whatever effect a verbal label may have on memory for form would seem to occur during the encoding phase of the memory task rather than upon the subsequent retrieval of the encoded information.

## REFERENCES

Clark, H. J. Recognition memory for random shapes as a function of complexity, association value, and delay. *Journal of Experimental Psychology*, 1965, **69**, 590–595.

Ellis, H. C. Transfer of stimulus predifferentiation to shape recognition and identification learning: Role of properties of verbal labels. *Journal of Experimental Psychology*, 1968, **78**, 401–409.

Ellis, H. C., & Daniel, T. C. Verbal processes in long-term stimulus recognition memory. *Journal of Experimental Psychology*, 1971, **90**, 18–26.

Ellis, H. C., & Homan, L. E. Implicit verbal responses and the transfer of stimulus predifferentiation. *Journal of Experimental Psychology* 1968, **76**, 486–489.

Ellis, H. C., & Muller, D. G. Transfer in perceptual learning following stimulus predifferentiation. *Journal of Experimental Psychology*, 1964, **68**, 388–395.

Price, R. H., & Slive, A. B. Verbal processes in shape recognition. *Journal of Experimental Psychology*, 1970, **83**, 373–379.

Vanderplas, J. M., & Garvin, E. A. The association value of random shapes. *Journal of Experimental Psychology*, 1959, **57**, 147–154.

(Received July 24, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 90–94

# BACKGROUND, SYMBOLIC, AND CLASS SHIFT IN SHORT-TERM VERBAL MEMORY[1]

DONALD B. REUTENER[2]

*Ohio State University*

The *S*s were given four trials using the "release from proactive inhibition" in short-term memory technique with shifts in class of item, background of item, and symbolic presentation of the item on Trial 4. For every type of shift, except the background shift with words, there was a significant increase in retention on the shift trial. A class shift led to greater increase in retention than a symbolic shift, which, in turn, was greater than a background shift. The data indicate that shifting a physical parameter of the stimulus is effective in producing a release from proactive interference in short-term memory and that increased retention on the shift trial can be explained in terms of memory processes rather than an increase in perception or an increase in learning due to novelty.

Melton (1963) has described a memory trace as undergoing three stages of transformation: (*a*) It is formed at the time of initial presentation; (*b*) it is stored according to how it was encoded by *S* as either the same or different from preceding information; (*c*) it is used or recalled, the utilization or recall depending on how it was stored. Recent experiments in short-term memory (STM) using the Peterson and Peterson (1959) design and incorporating a shift trial (shifting from one class of material to another) indicate that if the trace has been encoded as being of a similar class to preceding material, there will be interference at time of recall. If the trace has been encoded as being of a different class from the preceding material, there will be no, or less interference at time of recall. This phenomenon, referred to as release from proactive interference (PI) has been reviewed extensively by Wickens (1970).

The class shifts in these studies vary from one experiment to another, yet almost any type of class shift seems to lead to an increase in retention on the criterion trial. Some of the increases in retention on the criterion trial are great enough that level of recall on the shift trial matches the level of recall on the initial trial, while some of the increases in retention on the criterion trial reach only a proportion of the level of recall on the initial trial. The differences in the magnitude of the effect on the shift trial between different types of materials raises the possibility that something in addition to a release from PI is operating when performance in experimental groups is found to increase following a shift in material on the criterion item. Increased performance on the criterion trial may not all be due to the operation of memory processes during retention and recall. If the information on a shift trial is differently encoded from the preceding information, it is more likely to be retained as a function of its uniqueness. It is also possible that the uniqueness or novelty of the shift trial causes the memory trace for that information to become distinctive and form a new class even if the information presented belongs to the same formal class.

The purpose of the present study is to determine whether or not improvement in retention can be obtained on a shift trial through the manipulation of variables other than the class of items presented. A series of related experiments has been designed to test for (*a*) the effect of shifting the color of the background on which the to-be-remembered material is presented, (*b*) the effect of shifting the symbolic presentation of the material without a shift

---

[1] This research was supported in part by National Institute of Mental Health Training Grant MH08526.

[2] Requests for reprints should be sent to Donald B. Reutener who is now at Smith College, Department of Psychology, Northampton, Massachusetts 01060.

in meaning, and (*c*) the relative strengths of these shifts in comparison to a shift in the class of the to-be-remembered material. If performance increases on the shift trial with changes in the physical parameters of the stimulus, and if these performance increases are of the same magnitude as those observed for class shifts, some modification of the explanations of the PI release phenomenon may be necessary.

## METHOD

*Subjects.*—The *S*s were 432 male and female general psychology students at Ohio State University. The *S*s were run individually for one experimental session. Each *S* was randomly assigned to 1 of 18 groups with the restriction that 24 *S*s were assigned to each group.

*Lists and design.*—Eighteen lists were constructed wherein the materials to be learned and recalled were spelled out numbers (SpNu), arabic numbers (ArNu) or words (W). The SpNu's and ArNu's were selected so that no number appeared twice in a series of four trigrams and were drawn from the numbers 1 through 12. The words were taken from Category 14 of Battig & Montague (1969) and equated for frequency of usage according to the Thorndike-Lorge word count. Each word appeared only once in a series of four trigrams. Both number and word trigrams were rotated across trials so that each trigram appeared in each of four positions an equal number of times. Both the number and word trigrams had two forms: (*a*) white numerals or letters on a black background and (*b*) black numerals or letters on a white background.

The design consists then of 12 major experimental groups that shift background, symbolic presentation, class, and a combination of class and background on the fourth trial, and 6 major control groups that had no shift in material over four trials.

*Apparatus.*—The *S*s were seated in a sound-deadened room throughout the experiment. A Kodak Carousel projector was programmed by a Gerbrands tape timer to present to *S*s slides of the word or number trigrams and the cues which separated and paced the events. All slides were projected on a screen suspended from a wall approximately 4 ft. in front of the *S*s.

*Procedure.*—The procedure was that of the standard rehearsal preventive type Peterson and Peterson (1959) experiment. After *S* was seated in the chair, he was read instructions which described the entire procedure. When the trial began, an asterisk appeared on the screen for 1.5 sec. as a ready signal. This was followed by a word or number trigram which also appeared for 1.5 sec. and which *S* spoke aloud. Next in succession was the rehearsal preventive task slide, consisting of Stroop items obtained from Underwood (1966, p. 540). The *S* was instructed to name off the colors in which the words were printed as rapidly and as accurately as possible; this slide remained on the screen for 19.5 sec. A slide containing a question mark then appeared on the screen for 5.5 sec. as a signal to recall the previously presented trigram. The next trial was again indicated by the appearance of an asterisk. Projector change time was .5 sec. Thus, a single trial consisted of a 2-sec. ready period, a 2-sec. presentation of a trigram, a 20-sec. retention interval, and 6 sec. for recall of the trigram, for a total of 30 sec.

A recall was counted as correct if it was given during the appropriate interval. The *S*s were given one point for each item of the trigram recalled, regardless of the order in which they were given. The number of correct words in colors named by *S*s during the retention interval were also recorded. After completion of the experiments, *S*s were asked if they noticed any difference or differences between Trials 3 and 4.

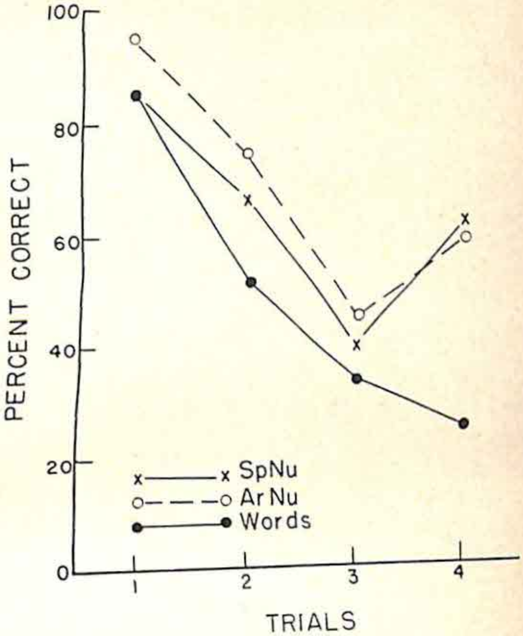


FIG. 1. Percentage of correct responses for experimental background shift groups.

## RESULTS

A test for the significance of the difference between two means for correlated samples showed that Trial 1 differed significantly from Trial 3 in all control groups as well as in all experimental groups ($p < .001$ for all cases), thereby replicating the previous findings of a severe decrement in recall over early trials in the typical STM paradigm.

Inspection of Fig. 1 suggests that shifting

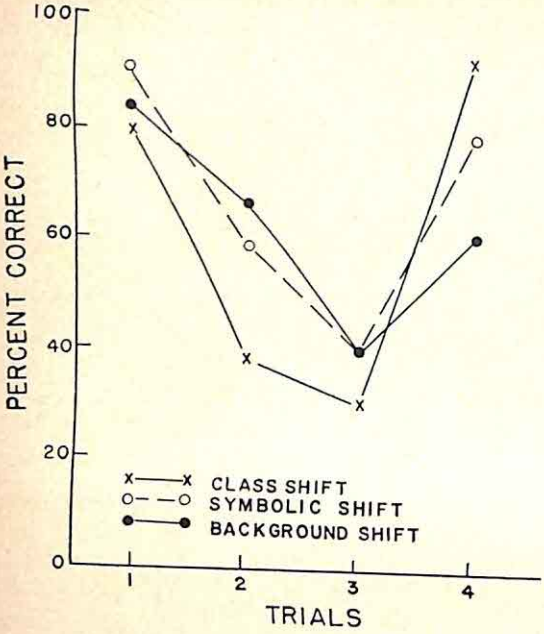


FIG. 2. Percentage of correct responses for combined class shift groups, combined symbolic shift groups, and combined background shift groups.

from a white background on the first three trials to a black background on the fourth trial, or the reverse, when the to-be-remembered items are W's had no effect on retention of items following the shift. Figure 1 also indicates that there is an increment in retention on the shift trial for the experimental groups with a background shift from white to black, or the reverse, when the to-be-remembered items are ArNu's or SpNu's.

Separate analyses of variance for the individual groups showed that the change in performance between Trials 3 and 4 for the experimental group recalling words was not significantly different, $F\ (1, 188) = .53$, $p > .20$, from the change in performance between Trials 3 and 4 for the appropriate control.

Similar comparisons showed that the changes in performance between Trials 3 and 4 for the experimental and control groups recalling ArNu's was significantly different, $F\ (1, 188) = 9.87$, $p < .005$, as was the change in performance for the groups recalling SpNu's, $F\ (1, 188) = 5.16$, $p < .05$. These results indicate a significant improvement on the fourth trial for the ArNu and SpNu experimental groups shifted from a white to a black background, or the reverse, on the fourth trial.

With a shift in the representational form of the stimulus from ArNu's to SpNu's or the reverse, performance on the fourth trial in both experimental groups is higher than for their appropriate controls. In both cases, the change in performance from Trial 3 to 4 between the experimental and control groups was significant beyond the .05 level.

Class shifts from words to either ArNu's or SpNu's and combination class and background shifts, where the experimental groups are shifted from words on a white background to either ArNu's or SpNu's on a black background all showed large increments in retention on the fourth trial. In all cases, the change in performance from Trial 3 to 4 between the experimental and control groups was significant beyond the .001 level. The form of these curves is very similar to that of the combined class shift groups in Fig. 2.

The overall change in retention on shift trials was obtained for each group using the method outlined in Wickens (1970). Preliminary tests indicated that the experimental groups switched from ArNu's to SpNu's or the reverse; SpNu's on a white background to SpNu's on a black background or the reverse; and W's to SpNu's or ArNu's were not significantly different, $p > .20$ in all cases; therefore the data from the three combinations above was pooled for further analyses. Figure 2 indicates for the above combined groups that increase in retention on the fourth trial is greatest for the class shift groups, next greatest for the symbolic shift groups, and least for the background shift groups. Analysis of variance showed that the change in performance between Trials 3 and 4 for the combined symbolic shift groups was significantly different, $F\ (1, 188) = 3.97$, $p < .05$, from that of the combined background shift group and the change in performance between Trials 3 and 4 for the combined class shift group was significantly different, $F\ (1,\ 188) = 12.94$, $p$

$< .001$, from that of the combined symbolic shift group.

Analysis of *S*s' performance on the distractor task showed that correct identification of the color words changed very little from the second through the fourth trial. All the groups averaged approximately the same number of correct responses on Trials 3 and 4. No group average changed by as much as plus or minus one correct response from Trial 3 to 4.

## DISCUSSION

On background shift trials, there were consistent differences between experimental and control groups for those *S*s receiving either ArNu's or SpNu's, but the experimental group receiving W's was not different from its appropriate control. With numbers as the to-be-remembered items it appears that the background is used as a codable dimension, or perhaps as a retrieval cue, whereas when words are the to-be-remembered items, background is not an important codable dimension or retrieval cue.

The increased retention on the background shift trial for the number groups may be interpreted as a function of increased novelty, or attention, and therefore increasing learning, aroused on the shift trial. This explanation seems unsatisfactory for two reasons: the background shift did not lead to increased retention for the background shift with words group, and in two number trigram groups, both of which involved a background shift from black to white, only two *S*s in each group could report that the shift trial was different in any respect from the three preceding trials. Increased retention on the shift trial in both of these groups was observed to be equal to that of the increase in groups which involved a shift from white to black background where the difference on the shift trial was reported by 16 and 19 *S*s respectively. If novelty or increased attention were factors leading to increased performance, it seems reasonable to assume that noticed changes would be more effective than nonnoticed. This, however, was not the case.

Another possible interpretation is that the words were encoded into a greater number of dimensions or category sets than the numbers. The shift in background for the number groups could lead to a new encodable category set, whereas in the word groups, the number of category sets already used being large, it is not. The background shift thus acts as a salient category for the number groups but not the word groups. Wickens[3] (personal communication, May 1970) has recently found that shifting from upper- to lowercase type or vice versa when the to-be-remembered material consists of noun or verb triads is not effective in producing the release from PI phenomenon. Turvey and Egan (1969), on the other hand, have observed the PI release phenomenon when shifting the area surrounding consonant trigrams. These data also seem to fit the above interpretation. A shift in a physical parameter of the stimulus having an effect when the to-be-remembered items are not words.

The class shift data confirm many earlier studies. The same dramatic increase in retention is observed in this study as in the studies reviewed by Wickens (1970), which involved either a CCC to NNN, or a taxonomic category shift. The effect of a combined class and background shift appears not to be additive in nature. Increase in retention was just as great for a class shift as for the combined class and background shift. This is most likely due to the great increase in the class shift groups where the retention on the shift trial is as high or higher than retention on the initial trial, leaving little room for improvement due to the addition of a background change.

The increase in retention for the symbolic representation shift group falls somewhere between the class and background shift groups. The shift in symbolic presentation might then be a more salient encoding dimension than the background shift. It is also possible that the great similarity in meaning between ArNu and SpNu allows interference to continue in category sets which the two types of numbers share, whereas this interference is not present when taxonomic or class shifts are introduced, because all new category sets in which no interference is generated are used by *S*s.

Overall, the effect of shifting a physical parameter of the stimulus in a STM task is to facilitate performance on the shift trial. Shifting the symbolic presentation does not lead, however, to as great an increase in retention as shifting the class of to-be-remembered material. Shifting the background does not lead to as great an increase in retention as a shift in symbolic presentation and is effective only when the to-be-remembered items are numbers. At the present time it appears as if these data lend support to a release from PI interpretation. The other interpretation of the shift trial increase in performance, heightened performance due to the novelty of the trigram, does not seem to be as cogent an explanation as that of PI release, as evidenced by the fact that noticed changes were no more effective in increasing retention than non-noticed changes.

---

[3] This study was made possible by National Institute of Mental Health Grant QEG-5-9-45027b-0053.

## REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

MELTON, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **54**, 157–173.

TURVEY, M. T., & EGAN, J. Contextual changes and release from proactive interference in short-term verbal memory. *Journal of Experimental Psychology*, 1969, **81**, 396–397.

UNDERWOOD, B. J. *Experimental psychology*. (2nd ed.) New York: Appleton-Century-Crofts, 1966.

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **1**, 1–15.

(Received July 31, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 95–99

# EFFECT OF NUMBER OF RESPONSE CATEGORIES ON DIMENSION SELECTION, PAIRED-ASSOCIATE LEARNING, AND COMPLETE LEARNING IN A CONJUNCTIVE CONCEPT IDENTIFICATION TASK

WILLIAM J. THOMSON[1]

*Vanderbilt University*

Multiple-category conjunctive concept identification problems may be processed in two stages, dimension selection followed by paired-associate learning. The effect of manipulating the number of response categories (four, six, or nine) on dimension selection, paired-associate learning, and complete learning was investigated in a 3 × 3 design involving 81 human *S*s. The number of response categories did not affect dimension selection, but a linear increase in total errors as a function of increasing response categories was found for paired-associate learning and complete learning groups. The results support a duoprocess theory of concept identification which assumes semi-independent and sequential dimension selection and paired-associate learning.

Two recent theoretical approaches specific to multiple-category conjunctive concept identification (CI) have been described. Trabasso and Bower (1964) propose a theory in which solution to a conjunctive problem is achieved by the independent solution of two subproblems. It is assumed that each subproblem is solved independently and concurrently, with a constant probability of solution on each trial. In contrast to the subproblem theory, Chumbley (1970) has proposed a duoprocess theory based on the separation of complete learning (CL) into two stages, dimension selection (DS) followed by paired-associate learning (PAL). Such a distinction had been previously suggested by Richardson and Bergum (1954), and mathematical models for simple (one relevant dimension) CI embodying DS and PAL have been developed by Bower and Trabasso (1964) and Reeve, Polson, and Dunham (1970). Chumbley, in considering a multiple-category conjunctive task, proposed that DS consists of picking values from exactly two dimensions to form a locally consistent hypothesis, i.e., a hypothesis consistent with the current stimulus. Following DS and the concomitant memorization of the initial hypothesis (e.g., red rectangle-1), *S* learns the remainder of the categorizations in a paired-associate fashion. Details of the model were tentative and will not be described, except to note that the PAL process is essentially identical to Bower's (1961) PAL model.

The purpose of the present experiment is to investigate the duoprocess approach by manipulating the number of response categories in multiple-category conjunctive CI. If the duoprocess model is valid, several results should obtain.

First, DS should be relatively unaffected by manipulating the number of response categories. It may be shown that a slight positive relationship is predicted between the number of response categories and trials in DS, since the probability of resampling decreases slightly as the number of response categories increases. However, relative to the similar positive relationship expected in the PAL stage, the increase in DS is minimal. Overstreet and Dunham (1969), using four and nine response categories in a conjunctive CI task, obtained a complex series of results which were interpreted as supporting the prediction that trials to criterion in DS were relatively unaffected by number of response categories. However, two problems arise in this interpretation. First, no feedback was given in the DS stage of learning, and thus comparisons with DS as a subset of

---

[1] Requests for reprints should be sent to William J. Thomson, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

CL (i.e., with response category feedback) must be qualified. Second, a trial was defined as the presentation of one stimulus from each response category. Hence, even though trials to criterion were virtually identical for the four- and nine-category groups, Group 9 saw more than twice as many stimuli. Thus the question remains open.

The second duoprocess prediction is that PAL should exhibit a linear increase in errors as a function of increasing the number of response categories (see Bower, 1961, for details). Also, if DS is unaffected by number of response categories, CL should likewise exhibit a linear increase in errors. Data on this latter point are equivocal. Walker and Bourne (1961) found an exponential increase in errors as response categories increased, but the increase in response categories was confounded with an increase in problem complexity (i.e., an increase in the number of relevant dimensions). Overstreet and Dunham (1969) found an increase in trials (defined above) to criterion as response categories increased in a PAL task, but total errors were not reported, while Polson and Dunham (1970) reported pilot data from a one relevant dimension CI task indicating a nonlinear, negatively accelerated function between errors and increasing response categories in a CL condition. Perhaps the clearest data come from a CL study by Kepros (1965), who varied the number of relevant dimensions (two, three, or four) in a conjunctive task, while independently varying the number of response categories (two, three, four, six, or eight). The major finding of interest to the present study was a linear increase in errors as a function of increasing response categories.

The final prediction concerns the sequential nature of the DS and PAL processes. If DS and PAL are independent and sequential, then performance in a CL task should be predictable from analyses of DS and PAL tasks. Evidence for equality in trials to criterion between summed DS and PAL, and a separate CL group, has been found by Richardson and Bergum (1954). However, the stimuli consisted of only three dimensions, and achievement of criterion in the DS stage was quite rapid, making the task for the PAL and CL groups virtually identical.

The present experiment was designed to provide further data concerning these three predictions from the duoprocess model. The effect of increasing the number of response categories on DS was tested by a series of groups with the task of discovering the relevant dimensions in the presence of response category feedback. The second prediction of a linear increase in errors and the third prediction of independent and sequential DS and PAL processes were investigated with two series of groups, denoted PAL and CL. Each of the PAL and CL conditions contained groups varying factorially on the number of response categories, and PAL and CL differed only in that PAL *S*s were told the relevant dimensions at the outset of the experiment.

## METHOD

*Subjects and stimuli.*—The *S*s were 81 introductory psychology students participating in partial fulfillment of a course requirement. Seven males and 2 females were randomly assigned to each of the nine experimental conditions.

Stimuli were composed of geometric forms consisting of four ternary dimensions: size (small, medium, large), color (red, blue, green), shape (square, circle, hexagon), and number of figures (one, two, three). In conditions where binary dimensions were used, only the first two of the above three values were present. Stimuli were mounted on 5 × 7 white cards for presentation.

*Procedure.*—Three instruction conditions DS, PAL, and CL were used, with three groups differing in number of response categories (four, six, or nine) in each condition. Thus a total of nine experimental groups were constructed: DS-4, DS-6, DS-9, PAL-4, PAL-6, PAL-9, CL-4, CL-6, and CL-9.

All *S*s were given detailed instructions indicating that the experiment was a straightforward classification task. Each of the four dimensions was described by showing *S* two cards containing two different values from each dimension. The names (but not the values) of each dimension were available to *S* throughout the experiment.

The conjunctive classification rule was described as a rule in which each category contained a unique joint occurrence of particular values from the two relevant dimensions (e.g., small, red objects are in Category B). The use of the conjunctive rule in sorting cards into different categories was illustrated by three examples. The first example used letters as dimensions; the second used parts of speech; and the third, geometric forms. Emphasis was placed on understanding the conjunctive rule, and *S*s were told that the same rule, based on two of the four dimensions, would be used in the actual experiment. Next, all *S*s were told the available response categories (a subset of the letters A through I). The particular responses for each *S* were available throughout the experiment.

The *S*s in Groups CL and PAL were told to indicate the classification of each card by writing the name of the response category on a response sheet. In addition, *S*s in the PAL group were told which dimensions were relevant (always color and size). DS *S*s were required to respond to each stimulus by writing the two dimensions upon which they felt the classification was based. Following each response, *S*s in all groups were told the correct response category for the card and were allowed to examine the card for a few seconds. Presentation was essentially self-paced. One to five *S*s were run simultaneously.

A set of 36 stimulus cards was constructed for each experimental group. In order to keep the number of stimuli and potential hypotheses identical for all groups, two of the dimensions were binary and the remaining two were ternary. Thus for the four-category groups, the relevant dimensions were binary; in the nine-category groups both relevant dimensions were ternary; and in the six-category groups, one relevant dimension was binary, and one was ternary. Detailed information about the number of values per dimension could have aided in problem solution, so such information was withheld in the instructions. Extensive debriefing indicated that only one *S* (in Group DS-6) may have used such information to solve the problem.

A trial was defined as the presentation of a single stimulus, and cycle of trials was defined as the presentation of a stimulus from each available response category. The *S*s were presented cycles of trials, and stimuli within a cycle were presented in random order, with the exception that two stimuli from the same category could not appear consecutively across cycles.

The *S*s were run to a criterion of one correct cycle of trials. A maximum of 180 trials was given in Cond. CL, and 144 trials for Cond. PAL and DS. A total of four *S*s (two in Group CL-9, one each in Groups CL-4 and DS-6) failed to achieve a criterion run within the allotted number of trials and were replaced.

## RESULTS

*Condition DS.*—An error was defined as any choice of dimensions other than both relevant dimensions. Analyses of variance were performed on total errors (TE), trial of last error (TLE), and trial of first selection of the relevant dimensions (TFS). No significant effect due to number of response categories was found for any of these measures. Mean values across groups were 42.22 for TE, 49.15 for TLE, and 10.96 for TFS.

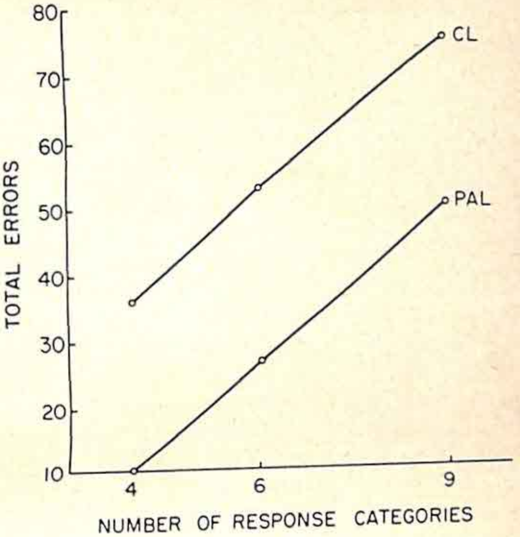


FIG. 1. Total errors as a function of number of response categories for paired-associate learning and complete learning conditions.

*Conditions CL and PAL.*—An analysis of variance for TE indicated main effects for number of response categories, $F\ (2, 48) = 14.2$, $p < .001$, and instructions, $F\ (1, 48) = 17.5$, $p < .001$, with the interaction nonsignificant. The means for the CL condition were 35.67, 52.78, and 75.22 for four, six, and nine response categories, respectively. Corresponding means for the PAL condition were 10.00, 27.11, and 49.89. An overall test for linear trend (Gaito, 1965) along response category indicated that 99.9% of the TE variance, $F\ (1, 48) = 28.4$, $p < .001$, was accounted for by the linear component (Fig. 1).

Similar main effects for response categories, $F = 15.79$, $p < .001$, and instructions, $F = 14.57$, $p < .001$, were found for TLE. For Cond. CL, the response category means ordered 51.44, 82.78, and 115.22; for PAL the means were 17.00, 48.22, and 79.33. Once again, significant linearity was found, $F\ (1, 48) = 31.20$, $p < .001$.

## Discussion

As predicted, DS was unaffected by manipulating the number of response categories. This result is consistent with the duoprocess approach, since dimension selection is assumed to be primarily influenced by changes in the number of available dimensions (e.g., Overstreet & Dunham, 1969).

The second prediction of a linear increase in TE for the PAL groups as a function of increasing response categories was also upheld. This result, in fact, was found in both CL and PAL groups, as indicated by the lack of an Instructions × Response Category interaction, and illustrated by the remarkable stability of the difference between CL and PAL for TE across response categories (i.e., 25.67, 25.67, and 25.33). Corresponding figures for differences between TLE across Groups 4, 6, and 9 were 34.44, 34.56, and 35.89. This constant difference between CL and PAL also lends support to the duoprocess notion that the only difference between CL and PAL is the addition of a dimension selection process in Group CL. Since dimension selection is independent of number of response categories, the difference between CL and PAL should be consistent.

Results related to the third prediction of independent and sequential DS and PAL processes are somewhat ambiguous. Consider TLE. As indicated above, a consistency of differences between CL and PAL is in line with independent and sequential processes. However, the size of this consistent difference (approximately 35 trials) should be mirrored in the DS group. Chumbley (1970) postulates that the initial selection of the two relevant dimensions initiates the PAL portion of the solution process. This assumption may be considered a theoretical lower bound for the number of trials in DS, and this bound is experimentally represented by TFS. Although the mean TFS was found to be consistent across response categories, the value ($\bar{X} = 10.96$) is considerably smaller than the difference between CL and PAL. On the other hand, the data also indicated that DS was not complete before PAL processing began. If DS had been complete, one would expect the mean TLE for DS ($\bar{X} = 49.15$) to approximate the 35-trial difference between CL and PAL. Such was clearly not the case.

At least two explanations for the lack of correspondance between the DS group and the difference between CL and PAL may be offered. First, the DS task may have been confusing to *S*, since he was required to give dimension names as responses, while receiving category feedback. If DS *S*s were confused, it is not unreasonable to assume that confusion would occur on some constant initial number of trials, bringing the actual number of nonconfusion trials more into line with the expected number of 35. In any case, the assumption that TFS is the critical event in initializing PAL processing would still be rejected.

A second explanation assumes that DS need not be complete before PAL processing begins. That is, *S* may in fact learn a constant small number of paired associates (e.g., one small red circle-B, one small blue square-F) and attempt to deduce the relevant dimensions. Thus DS, rather than random selection, is deductive. Such processing would still be independent of the number of response categories, but the total number of trials in the PAL phase would be reduced since a small number of PALs would be known when DS was complete. Thus the difference between CL and PAL would underestimate the number of trials taken by an independent DS group. Informal evidence, obtained through extensive debriefing of DS *S*s, indicated that a large proportion of the paired associates were known (even though such information was not requested in the task) and that *S*s did attempt to deduce the relevant dimensions, rather than randomly sample.

In general, the present results lend strong support to the duoprocess notion. DS was found to be uninfluenced by manipulating the number of response categories, even though some questions remain concerning the independent and sequential nature of DS and PAL. In addition, the predicted linear increase in total errors with increasing response categories was found for both Cond. PAL and CL.

## REFERENCES

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, *26*, 255–280.

Bower, G. H. & Trabasso T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964.

Chumbley, J. I. *A duoprocess theory of concept learning*. (Tech. Rep. 70–3) Amherst, Mass.: Cognitive Processes Laboratory, University of Massachusetts, 1970.

Gaito, J. Unequal intervals and unequal *n* in trend analyses. *Psychological Bulletin*, 1965, *63*, 125–127.

Kepros, P. G. Identification of conjunctive concepts as a function of stimulus and response complexity. Unpublished doctor's thesis, University of Utah, 1965.

OVERSTREET, J. L., & DUNHAM, J. L. Effect of number of values and irrelevant dimensions on dimension selection and associative learning in a multiple concept problem. *Journal of Experimental Psychology*, 1969, **79**, 265–268.

POLSON, P. G., & DUNHAM, J. L. A comparison of two types of theories of multiple-category concept identification. Paper presented at the Midwestern Mathematical Psychology meetings, Bloomington, Indiana, April 1970.

REEVE, M. R., POLSON, P. G., & DUNHAM, J. L. The size of focus samples in multiple-category concept identification. *Psychonomic Science*, 1970, 20, 125–126.

RICHARDSON, J., & BERGUM, B. O. Distributed practice and rote learning in concept formation. *Journal of Experimental Psychology*, 1954, **47**, 442–446.

TRABASSO, T., & BOWER, G. H. Component learning in the four-category concept problem. *Journal of Mathematical Psychology*, 1964, 1, 143–169.

WALKER, C. M., & BOURNE, L. E., JR. Concept identification as a function of amounts of relevant and irrelevant information. *American Journal of Psychology*, 1961, 74, 410–417.

(Received August 2, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 100–104

# NONREINFORCED TRIALS IN CONCEPT IDENTIFICATION: PRESOLUTION STATISTICS AND LOCAL CONSISTENCY[1]

LEONA S. AIKEN[2]
*Temple University*

JOHN L. SANTA
*Douglass College, Rutgers University*

ALAN B. RUSKIN
*Purdue University*

The effects of nonreinforced (blank) trials during presolution of a concept identification task were examined. A partial reinforcement group received 50% randomly administered blank trials; a control group was reinforced on all trials. Presolution statistics, number of called errors, and trials to trial of last error indicated that *S*s did not resample hypotheses following blank trials. The degree of local consistency, measured by matching of responses on Trial $n$ to actual or inferred feedback on Trial $n - 1$ was weaker following uncalled successes and errors than following both called successes and called errors ($p < .01$ in all cases). Consistency was greater after called successes than called errors ($p < .01$). These results are explained through a learning-performance distinction and the notion of a passive state.

Levine (1966) has employed a blank-trials procedure to evaluate *S*s' hypothesis behavior on concept problems where the solution depends on a single relevant dimension. The technique consists of embedding sets of nonreinforced trials within a test sequence. The technique allows Levine to infer *S*'s current active hypothesis.

One assumption central to the use of the blank trials procedure is that *S* does not change his current hypothesis on nonreinforced trials. A failure to resample hypotheses is also the distinguishing characteristic of success trials, according to the early quantitative models of concept identification (Bower & Trabasso, 1964; Restle, 1962; Trabasso & Bower, 1966). This raises the possibility that *S*s treat nonreinforced trials as if they were success trials in concept identification tasks. Levine (1966) has, in fact, provided support for the notion that his *S*s treat nonreinforced trials as if they were success trials by demonstrating that over 90% of his *S*s appear to maintain a hypothesis on both nonreinforced and success trials. However, Levine's task was considerably different from the conventional concept identification problem. The problems involved only four dimensions, *S*s were extremely practiced, and fewer than 20% of all responses were reinforced. Therefore, it is reasonable to ask, in general, whether nonreinforced trials are truly blank, i.e., equivalent to success trials.

Such a question can be asked at several levels. One approach is to manipulate the percentage of reinforced trials in a concept problem, and then to examine the effect on various learning statistics, such as the mean number of errors. At this level it is possible to ask if resampling occurs on nonreinforced trials. If nonreinforced trials are equivalent to success trials in terms of these learning statistics, one would predict, in accord with previous work (Trabasso & Bower, 1966), that the mean number of called errors should be independent of the percentage of reinforcement. Alternatively, if *S*s resample on some proportion of the nonreinforced trials, an increase in called errors would be expected.

Given this overall test of equivalence, a finer breakdown of the data might be employed. Falmagne (1970) examined *S*s' consistency curves by finding the probability of repeating on Trial $n$ the response

---

[1] This research was in part supported by Grant in Aid of Research 400-101-02 from Temple University to the first author and Rutgers University Research Council and Biomedical Support Grant to the second author.

[2] Request for reprints should be sent to L. S. Aiken, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

which was reinforced on Trial $n - 1$. This probability of repeating the reinforced outcome (tracking) was a linear function of the number of dimensions ($k$) which retained the same values from Trial $n - 1$ to $n$. The measure is essentially a measure of local consistency, or the extent to which *S*s retain and use the stimulus information from the previous trial. While tracking was evidenced, it was not complete, i.e., tracking was less than $\frac{k}{n}$. This fact was a major basis of Falmagne's argument for the existence of a passive state, in which *S*s respond randomly, without using a hypothesis generated from the previous trial to guide their action. The probability of entering the passive state on any given trial was assumed to depend on *S*'s state, response, and outcome of the previous trial.

In the present circumstance, one would suspect that reinforcement should influence the occurrence of a passive state, such that *S*s would be more prone to lapse into a passive state following a nonreinforced trial. To the extent that *S*s enter the passive state, the slope of the consistency curves should tend to be independent of $k$. It is therefore possible for differences between nonreinforced and success trials to arise in terms of local consistency, even in the absence of differences on simpler learning statistics.

The present study attempts to examine the differences between nonreinforced trials and success trials by varying the percentage of reinforcement. The data are then analyzed on the basis of both standard learning statistics and an examination of consistency curves.

## METHOD

*Stimuli.*—Stimuli were the 32 unique five-dimensional binary patterns formed from the following binary attributes: (a) fill (filled, empty), (*b*) size (large, small), (*c*) shape (square, circle), (*d*) orientation (45°, 135°), and number (three, two).

*Subjects.*—A total of 86 undergraduates participated in the experiment to satisfy course requirements. Fifty *S*s were from Purdue University, and the remaining 36 were from Douglass College, Rutgers University.[3]

*Experimental conditions.*—There were two conditions. In Cond. 1, the control condition (C), *S*s received feedback on all trials. In Cond. 2, the partial reinforcement condition (P), *S*s received feedback on 50% of all trials. The determination of feedback was random within each set of 32 trials, in which each different stimulus was presented only once.

*Procedure.*—The *S*s from both universities were assigned randomly to the two conditions, with an equal number of *S*s from each university assigned to each condition. They were instructed that they would view patterns which varied in fill, size, shape, orientation, and number. They were told that stimuli belonged to one of two categories, and that by responding to each stimulus and receiving feedback, they were to discover the category rule. The partial reinforcement group was also informed that on some trials they would receive no feedback, but were to continue responding when this occurred.

Stimuli were arranged in random permutations. A subset of these permutations constituted a protocol. Stimuli were projected for 6 sec. each. After 3 sec., *S*s were cued to respond verbally. Verbal feedback was given after the 3-sec. response period. Within each condition, the dimension chosen as relevant varied across *S*s. The criterion for solution was a minimum of 16 consecutive correct responses.

## RESULTS

Two learning statistics, the mean number of called errors (CE) and the mean trial of last error (TLE) were compared across the partial reinforcement and control groups. The mean CE for the partial reinforcement group was 6.74; for the control group, 5.75. These means were not significantly different, $t(81) = 1.02$, $p > .20$. The partial reinforcement group required approximately twice as many trials to last error as did the control group, with means of 27.57 and 12.90 for the two groups, respectively, $t(81) = 3.36$, $p < .01$.

These data point to the similarity of blank and success trials in terms of learning statistics. The equality of CE's across conditions suggests that *S*s did not resample hypotheses on nonreinforced trials.

*Local consistency.*—To further examine the effects of called versus uncalled trials, an analysis was made of local consistency on Trial $n$ as a function of trial type on

---

[3] The data of three *S*s, two from Purdue and one from Rutgers, were discarded due to failure to reach solution in 128 trials.

TABLE 1

MEAN PROPORTION CONSISTENT RESPONSES FOR EACH TRIAL TYPE

| Trial type | | No. of common dimensions | | | |
|---|---|---|---|---|---|
| | | One | Two | Three | Four |
| CS | p[a] | .31 | .39 | .49 | .67 |
| | N[b] | 100 | 70 | 174 | 101 |
| CE | p | .37 | .42 | .50 | .64 |
| | N | 71 | 128 | 125 | 61 |
| US | p | .44 | .44 | .51 | .64 |
| | N | 41 | 106 | 98 | 55 |
| UE | p | .42 | .52 | .54 | .63 |
| | N | 33 | 72 | 89 | 49 |

[a] Proportion consistent responses.

[b] Number of observations on which proportion is based.

Trial $n-1$. The protocol of each *S* was divided into Trial Pairs $n-1$ and $n$ having one through four dimensions in common. (Due to the use of random permutations of the stimuli, a repetition of a stimulus yielding a pair having five common dimensions did not occur. Pairs having zero common dimensions were not analyzed, due to the infrequency of their occurrence, i.e., in only 45 of the 1,418 available presolution pairs). These pairs at each common dimension level were then divided into two subjects for Cond. C, depending on the trial type, called error (CE) or called success (CS), which had occurred on Trial $n-1$. For the partial reinforcement Cond. *P*, pairs were divided into four subjects for the four trial types, CE, CS, UE (uncalled error), and US (uncalled success) which had occurred on Trial $n-1$. For each trial type and number of common dimensions, the number of responses to these trials which were consistent with feedback on Trial $n-1$ were counted. For uncalled trials, both UE and US were scored as if feedback had been "correct" on Trial $n-1$. This analysis is analagous to that performed by Falmagne (1970).

The partial reinforcement versus control conditions were tested for equivalence of local consistency following CS and CE trials. Two 2 × 4 unweighted means analyses of variance, one for each trial type, were performed. Factors were condition (C, P) and number of common dimensions (one, two, three, four). The dependent variable was the consistency of the response to each pair of trials, i.e., a pair was scored "one" for a consistent response, "zero" otherwise. For pairs containing a CS trial on Trial $n-1$, equal mean proportion consistent responding was found across conditions, $F(1, 546) = .00, p > .20$. Furthermore, no interaction was found between condition and number of common dimensions, $F(3, 546) = 1.63, p > .10$, indicating homogeneous consistency after CS's in the two conditions. A significant effect of number of common dimensions, $F(3, 546) = 13.43, p < .01$, reflected the monotonic increase in proportion consistent responses as a function of number of common dimensions (see Table 1).

Similar results were found for called errors. There was neither a condition main effect, $F(1, 337) = .37, p > .20$, nor a Condition × Number of Common Dimensions interaction $F(1, 337) = 1.81$, $p > .10$. The main effect of number of common dimensions was significant, $F(3, 337) = 4.53, p < .01$, again reflecting the monotonic increase in local consistency as a function of number of common dimensions (see Table 1).

Since no evidence of differential consistency in either CS or CE trials was found in the two conditions, data were pooled across conditions for further analyses of local consistency. The mean proportions of consistent responses after both uncalled successes and errors as a function of number of common dimensions are also shown in Table 1.

Weighted regression analyses were performed to determine the slope of the least-squares line, showing consistency as a function of number of common dimensions. Analyses were performed separately for each trial type. Weights for the analyses were the number of observations upon which each proportion was based, as given in Table 1. Results of the regression analyses are shown in Table 2. For all trial types, the best-fit line deviated from zero slope ($p < .01$ in all cases), indicating the existence of some degree of local consistency in each case. The consistency was not perfect, however, as slopes of unity would have been expected if perfect consistency had been obtained.

Since slopes varied as a function of trial types, various slopes were tested for significance of difference. Comparisons were made between slopes of consistency following CS versus CE, CE versus US, and US versus UE. To control for collective error in the set of tests, Scheffé critical values were employed. The slope following CS trials was found to be significantly steeper than that following CE trials, $t$ (935) = 22.91, $p < .01$. The same was found for CE versus US trials, $t$ (681) = 12.70, $p < .01$, and for US versus UE trials, $t$ (539) = 9.86, $p < .01$. In sum, greater consistency was found following success over error trials, and called over uncalled trials.

## Discussion

The analysis of learning statistics supported Levine's (1966) assumptions regarding nonreinforced trials. No significant differences were obtained between standard presolution statistics of the two conditions of the experiment, suggesting that *S*s did not resample hypotheses on nonreinforced trials.

A finer breakdown of the data in terms of consistency curves, conditionalized on feedback, revealed highly significant differences in the extent of tracking. In accord with Falmagne (1970), local consistency was greater following called successes than called errors, as shown by the difference in the slopes of the two consistency curves. The observed slope for called successes is quite comparable to Falmagne's data. Furthermore, the slope for called errors is less steep than that for called successes, as also found by Falmagne. The data of this study differ from those of Falmagne in one respect. She found an essentially flat consistency curve following called errors, which indicated that *S*s did not respond with regard to the previous stimulus on those trials on which they had made an error. Following errors, they apparently resampled hypotheses without using the stimulus information from the prior trial to do so. The present data are somewhat contradictory, in that evidence was found for partial local consistency following CE trials.

TABLE 2

Weighted Regression Analyses of Local Consistency as a Function of Number of Common Dimensions

| Trial type | Regression coefficient | $t^a$ | $df$ | $p$ |
|---|---|---|---|---|
| CS | .59 | 112.55 | 552 | <.01 |
| CE | .43 | 76.48 | 383 | <.01 |
| US | .35 | 40.09 | 298 | <.01 |
| UE | .28 | 31.40 | 241 | <.01 |

[a] Test for significance of difference from zero slope.

The data also exhibit some degree of tracking following nonreinforced trials. This suggests that *S*s did consider the stimulus which occurred on a blank trial in formulating their next response. However, both uncalled successes and uncalled errors produced consistency curves which were less steep than those of the called trials. This, in turn, suggests that it is the probability of consideration of the stimulus on Trial $n - 1$ which differs from the reinforced to blank trials.

Perhaps the simplest way to consider the present data is in terms of a learning-performance distinction. Nonreinforced trials have essentially no effect in terms of conventional learning statistics. The *S*s did not seem to resample hypotheses on uncalled trials. However, the finer breakdown of the data in terms of consistency curves revealed performance differences following nonreinforced trials. A possible interpretation of this learning-performance distinction can be attained through an elaboration of Falmagne's (1970) hypothesized passive state.

The passive state is one in which *S* temporarily responds by chance without the use of a hypothesis. In the partial reinforcement condition, it is quite likely that the absence of feedback will induce *S* to lapse into the passive response state. It should be pointed out that the passive state is one in which *S* responds haphazardly, but it does not imply *S* is failing to work on the problem. On the contrary, *S* is quite likely to lapse into the passive response state when he is working on the problem. The passive state is then simply a response, or performance, state, the occurrence of which has no consequence on the learning process per se. There would be essentially no reason for *S* to make use of such a state if the act of responding on the basis of a hypothesis did not require the employment of processing capabilities.

In the normal sequence of events, *S* is trying to keep track of previous information and current hypotheses. This process is continually being disrupted by both the requirement to respond on the basis of a hypothesis, and by the feedback information which must be processed. A nonreinforced trial provides *S* with an opportunity to concentrate on rehearsal of his hypothesis set. Therefore, in order to maintain an effective rehearsal set, *S* might have a tendency to produce a passive state response on the trial following nonreinforcement. Such a notion of an increased proclivity to a passive performance state would explain the difference in consistency following nonreinforcement, in spite of the equivalence of the two experimental conditions in terms of overall learning statistics.

## REFERENCES

Bower, G., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Standord University Press, 1964.

Falmagne, R. Construction of a hypothesis model for concept identification. *Journal of Mathematical Psychology*, 1970, **7**, 60–96.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

Trabasso, T., & Bower, G. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*, 1966, **3**, 163–173.

(Received August 3, 1971)

Journal of Experimental Psychology
1972, Vol. 93, No. 1, 105–110

# ENCODING AND IMMEDIATE SERIAL RECALL OF CONSONANT STRINGS[1]

BARRY H. KANTOWITZ[2]
*Purdue University*

PETER A. ORNSTEIN
*Princeton University*

MARIAN SCHWARTZ
*University of Wisconsin*

Immediate serial recall of consonant strings differing in digram frequency was tested for four encoding conditions: Pronouncing, Spelling, Grouping, and a No-Instruction Control condition. In all encoding conditions, recall was better for strings of high digram frequency than for strings of low digram frequency. For *S*s encountering strings in the Low–High order, conditions demanding overt vocalization (pronouncing or spelling) produced superior recall; in the High–Low order, no effect of encoding instructions was found. Significant interactions involving serial position indicated that the effects of digram frequency and of encoding condition were both concentrated in the last half of the serial position curve. In the absence of an interaction between digram frequency and encoding condition, it was concluded that the digram frequency effect was not mediated by encoding processes.

Recall of letter strings is greatly aided when sequential constraints are imposed upon the strings. Mayzner and Schoenberg (1964) demonstrated that the immediate serial recall of consonant strings composed of high-frequency letters was significantly better when the letters were arranged as high-frequency digrams (overlapping pairs of letters) than when the same letters were arranged as low-frequency digrams.

In interpreting this digram-frequency effect, Mayzner and Schoenberg defined digrams as verbal units and invoked the "spew hypothesis" of Underwood and Schulz (1960), which states that the availability of verbal units is directly related to the frequency with which the units have been experienced. The mechanism through which high digram frequency could increase availability remains unclear, however. One possibility is that high-frequency digrams are more available because *S*s are able to encode them more efficiently.

The major purpose of the present study was to determine whether the digram-frequency effect (i.e., the finding that strings composed of high-frequency digrams are better recalled than those composed of low-frequency digrams) could be attributed to encoding operations performed by *S*. Encoding was manipulated through instructions. It was reasoned that encoding operations would be implicated as a mechanism in the digram-frequency effect if instructions to encode had a differential effect upon recall of the two types of strings. Aside from the possible differential effects specified above, a second purpose of the present experiment was to determine the effects of various kinds of encoding instructions upon overall recall and upon the serial position function.

Three types of encoding instructions were used. In the Pronouncing condition, *S*s were instructed to pronounce each letter string aloud as a collection of syllables. In the Spelling condition, *S*s were instructed to spell each string aloud in groups of three letters. In the Grouping condition, *S*s were instructed to group the letters in each string, but were not instructed to pronounce or spell aloud. Overt vocalization, therefore, was involved in the first two of these conditions but not the third. Specific instructions to group were given only in the last two of these conditions; pronouncing as syllables, however, necessarily involves a kind of grouping, so that a grouping operation may be considered to be involved in all three of these conditions. There was also a No-Instruction (NI) control condition, in which no special encoding instructions were given.

---

[1] This research was supported in part by an XL Grant from the Purdue Research Foundation to the first author and by National Institute of Mental Health Grant MH 17620-01 to the second author. This report was written while the second author was a NATO Postdoctoral Fellow at University College, London. The authors are indebted to Pat Mallonee who collected the data. We also thank M. S. Mayzner for providing a copy of his stimulus material.

[2] Requests for reprints should be sent to Barry H. Kantowitz, Purdue University, Department of Psychology, Lafayette, Indiana 47907.

If the mechanism underlying the digram-frequency effect involves the encoding operations of pronouncing, spelling, or grouping, this should be reflected in an interaction between the variables of encoding condition and digram frequency. For example, it might be hypothesized that the digram-frequency effect results when *S*s adopt the strategy of encoding each letter string into a collection of syllables. Such a hypothesis would stem from the finding (Laughery & Pinkus, 1968) that immediate serial recall of meaningless letter strings is higher when the letter sequences are more pronounceable, combined with the assumption that consonant strings of high digram frequency would yield more pronounceable syllables than those of low digram frequency. If this hypothesis is correct, then the digram-frequency effect should be greater in the Pronouncing condition than in the NI control and should be reduced or absent in the Spelling and Grouping conditions, since the requirement to use other strategies should minimize the tendency to encode the strings into syllables.

Apart from the possibility that encoding instructions would have differential effects upon the two types of strings, there are grounds for predicting that the various encoding operations would increase the overall level of recall. Relative to an uninstructed control, the instruction to spell letters aloud during presentation has been found to facilitate serial recall of consonant strings (Murray, 1966, 1967); and, for various types of stimuli, instructions to group the stimuli into patterns without overt vocalization have markedly facilitated recall (Palmer & Ornstein, 1971; Ryan, 1969; Wickelgren, 1967). Accordingly, it would be expected that, regardless of type of string, recall in the Spelling and Grouping conditions should be superior to that in the NI control. In two experiments which lacked a no-instruction control (Lindley, 1966; Stark & Calfee, 1970), instructions to pronounce trigram sets or letter strings produced better recall than instructions to use other kinds of encoding operations. Such results suggest that recall in the Pronouncing condition would probably surpass that in the NI control, although no firm prediction can be made.

A third purpose of the present study was the evaluation of possible order effects. With type of string (high digram frequency and low digram frequency) as a within-*S* variable, the encoding and rehearsal strategies adopted by *S* may be influenced by the order in which the two types of strings are presented. In the Mayzner and Schoenberg (1964) study, the different types of strings were presented in a randomly intermixed order, which precluded investigation of a potential order effect. The present study used a blocked presentation, which allowed the evaluation of this effect.

## METHOD

*Stimuli.*—The stimulus materials in the test series were 20 nine-letter strings. Each string consisted of some permutation of the nine consonants occurring most frequently in English (C, D, F, H, L, N, R, S, T). There were 10 High strings (high digram frequency: e.g., RCHNFSTLD) and 10 Low strings (low digram frequency: e.g., SNRCFHLDT). The summed digram frequency for the eight overlapping digrams in each string (e.g., RC, CH, HN, etc.) averaged 773 for the High and 114 for the Low strings, based upon the digram frequencies for written English given by Underwood and Schulz (1960). The following additional restrictions were placed on the construction of the strings: (*a*) Within the set of 10 High strings, and also within the set of 10 Low strings, each letter occurred in at least seven of the nine serial positions, including the first and last position, and no individual digram occurred in more than 5 of the 10 strings. (*b*) The number of digram duplications was equated between the High and Low strings; i.e., for every digram that occurred $N$ times among the High strings, there was a corresponding digram that occurred the same $N$ times among the Low strings. These added restrictions, although necessarily diminishing the mean difference in digram frequency between the High and Low sets, allowed greater control over position and repetition factors which might have been confounded with digram frequency in the Mayzner and Schoenberg (1964) study. Each string was typed in capital letters on a separate slide. There were no spaces between successive letters.

*Procedure.*—A Carousel projector was used to present each slide for 6 sec. Immediately after each 6-sec. presentation, *S* was allowed 10 sec. for written recall. A card with nine spaces was provided for entry of each string. The *Ss* were instructed to write the letters in the spaces corresponding to their correct serial order, but the instructions did not specify the temporal order in which the letters were to be entered. The *Ss* were encouraged to guess when necessary. At the end of the 10-sec. recall period, the projector flashed a white light which signaled *S* to stop writing and turn over the card. After an interval of 5.5 sec., the next slide was projected.

Along with the general instructions, each *S* was shown an illustrative slide (PWJGMBZVK) together with instructions specific to the encoding condition. In the Pronouncing condition, *E* indicated that the letters on the slide could be pronounced as the following set of syllables: PWIJ-GIM-BIZVIK. In the Spelling condition, *E* indicated that the slide could be spelled aloud letter by letter and spelled out the sample string with a pause after every third letter: PWJ GMB ZVK. In the Grouping condition, *E* indicated that the slide could be grouped into three-letter segments and held up a sheet lettered PWJ GMB ZVK, but did not pronounce or spell the letters. In each of these three conditions, *Ss* were instructed to follow the method illustrated with the sample slide for all subsequent slides. Those *Ss* in the first two of these conditions were monitored to ensure they pronounced or spelled each slide aloud as instructed; one *S* who failed to follow instructions in the Pronounce condition was replaced. In the NI control condition, *Ss* were merely shown the illustrative slide without any special encoding instructions. The first illustrative slide was followed by a second illustrative slide and then by eight practice slides, each consisting of some permutation of the illustrative-slide letters.

On the test slides, all *Ss* were shown the same 20 strings. A randomly designated half of the 20 *Ss* in each encoding group received the 10 strings of low digram frequency followed by the 10 strings of high digram frequency (Low–High block order), with the remaining 10 *Ss* receiving the opposite order of presentation (High–Low block order). A 10 × 10 Latin square was used to counterbalance order of presentation of the 10 strings within each block.

*Subjects.*—Eighty women undergraduates at the University of Wisconsin participated in the experiment, fulfilling a course requirement in introductory psychology. Each *S* was tested individually. Twenty *Ss* were randomly assigned to each of the four encoding conditions.

## RESULTS

*Correct responses.*—A correct response was defined as a correct letter recorded by *S* in its proper serial position. Since there were nine letters per string, a maximum recall score of nine was achieved if a string had been perfectly reproduced. The mean number of correct responses per string, as a function of the various encoding conditions, is presented in Table 1. As can be seen, retention was better for the High strings than for the Low. A four-factor mixed-design analysis of variance with two between-*S* factors (encoding instructions and block order) and two within-*S* factors (digram frequency and trials), revealed this digram-frequency effect to be highly significant, $F\ (1, 72) = 84.12$, $p < .001$. There was no significant effect of encoding instructions, $F\ (3, 72) = 1.91$, $p > .05$, and no interaction between digram frequency and encoding instructions, $F\ (3, 72) = .48$, $p > .05$. Furthermore, there was neither a significant trials effect, $F\ (9, 648) = .46$, $p > .05$, nor any significant interaction between trials and any other variable.

TABLE 1

MEAN CORRECT RESPONSES PER ITEM

| Digram frequency | Encoding cond. | | | |
|---|---|---|---|---|
| | NI control | Grouping | Spelling | Pronouncing |
| Low | 5.13 | 5.46 | 5.75 | 5.75 |
| High | 6.08 | 6.26 | 6.43 | 6.68 |

Table 2 shows the mean correct responses per string for each block order (Low–High and High–Low). The interaction of encoding instruction and block order which is apparent in Table 2 was significant,

TABLE 2

MEAN CORRECT RESPONSES PER ITEM

| Block order | Block | Encoding cond. | | | |
|---|---|---|---|---|---|
| | | NI control | Grouping | Spelling | Pronouncing |
| Low–High | 1 | 4.49 | 5.19 | 6.06 | 6.17 |
| | 2 | 5.62 | 6.29 | 6.62 | 6.82 |
| | $\bar{X}$ | 5.06 | 5.74 | 6.34 | 6.50 |
| High–Low | 1 | 6.54 | 6.23 | 6.23 | 6.54 |
| | 2 | 5.76 | 5.72 | 5.43 | 5.33 |
| | $\bar{X}$ | 6.15 | 5.98 | 5.83 | 5.94 |

$F\ (3, 72) = 3.98$, $p < .025$. Comparisons among the four encoding instruction means within the Low–High block order by Duncan multiple-range test revealed performance in both the Spelling and the Pronouncing conditions to be significantly ($p < .05$) superior to that of the NI control condition. The corresponding analysis for the High–Low block order revealed no significant differences, ($p > .05$ for all comparisons). Indeed, a more liberal test than the Duncan multiple-range, two-tailed $t$ tests, also failed to show any significant differences as a function of encoding instructions within the High–Low block order. It appears, therefore, that the instructions to group without overt vocalization had no significant effect upon performance, whereas the two encoding conditions which required overt vocalization—spelling and pronouncing aloud—had a significant effect, but only for those *S*s receiving the Low strings first.

Additional $t$ tests comparing the two block orders for each encoding condition revealed that for the NI control condition, performance was better in the High–Low than in the Low–High block order, $t\ (72) = 2.81$, $p < .01$. For the remaining three encoding instructions, no significant differences were found between the two block orders.

An additional four-factor mixed-design analysis of variance was performed, with the same two between-*S* factors, with trials as one within-*S* factor and blocks (first or second, regardless of whether the block contained High or Low strings) replacing digram frequency as the second within-*S* factor. The critical finding was that the three-way interaction between encoding instructions, block order (Low–High or High–Low), and block (first or second) failed to reach significance, $F\ (3, 72) = .47$, $p > .05$, indicating that the relation between encoding instructions and block order was constant over both halves of the experiment. That is, as can be seen in Table 2, among those *S*s who first encountered strings of high digram frequency, there was no significant effect of encoding condition, either in the first block (High) or in the second block (Low). On the other hand, those *S*s who first encountered strings of low digram frequency were influenced by those encoding instructions which called for overt vocalization, and those instructions were equally facilitating in both the first block (Low) and the second block (High).

*Serial position curves.*—For all *S*s, the number of correct responses in each of the nine serial positions were summed over each of the two 10-trial blocks. The mean percentage of correct responses expressed as a function of serial position may be found in Fig. 1 and 2. A four-way mixed-design analysis of variance with two between-*S* factors (encoding instructions and block order) and two within-*S* factors (serial position and digram frequency) revealed a highly significant effect of serial position, $F\ (8, 576) = 170.41$, $p < .001$. A highly significant interaction between digram frequency and serial position, $F\ (8, 576) = 4.28$, $p < .001$, is shown in Fig. 1. There is no digram-frequency effect in the first serial position, but the near-ceiling performance level at that position may have precluded any difference between the two types of strings; or it may be that since the first serial position is not preceded by any letter, digram frequency is not a meaningful variable when the first serial position is considered. Moreover, ceiling effects do not appear at later serial posi-

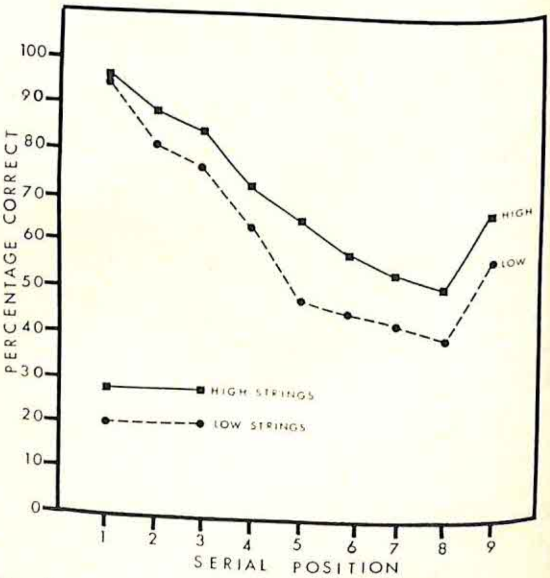


FIG. 1. Percentage of correct responses as a function of serial position and digram frequency.

tions. Of the 80 *S*s tested, only 25 exhibited perfect scores for High strings at Serial Position 2, 22 at Serial Position 3, and 9 at Serial Position 4; for Low strings 12, 9, and 5 *S*s recalled perfectly at these serial positions. As serial position increases, the difference between High and Low strings increases monotonically until the fifth serial position after which it decreases nonmonotonically but remains greater than for the first four serial positions. Hence, the relative superiority of the High strings can be localized largely in the last half of the serial position function.

A highly significant Serial Position × Encoding Instructions interaction, $F$ (24, 576) = 5.72, $p < .001$, is shown in Fig. 2. Again it is the last half of the serial position curve which distinguishes among conditions. Recency effects were found in all conditions, but were more marked in the two conditions involving overt vocalization (Spelling and Pronouncing) than in the other two. The three-way interaction among digram frequency, encoding instructions, and serial position did not reach significance, $F$ (24, 576) = .55, $p > .05$, indicating that the effect of digram frequency upon the shape of the serial position curve was the same for all encoding conditions. The serial position curves did not differ for the two block orders, $F$ (8, 576) = 1.14, $p > .05$. A significant three-way interaction among serial position, digram frequency, and block order, $F$ (8, 576) = 3.23, $p < .005$, indicated a greater recency effect for the second test block for both High and Low strings. All remaining interactions with serial position failed to reach the .05 level of significance.

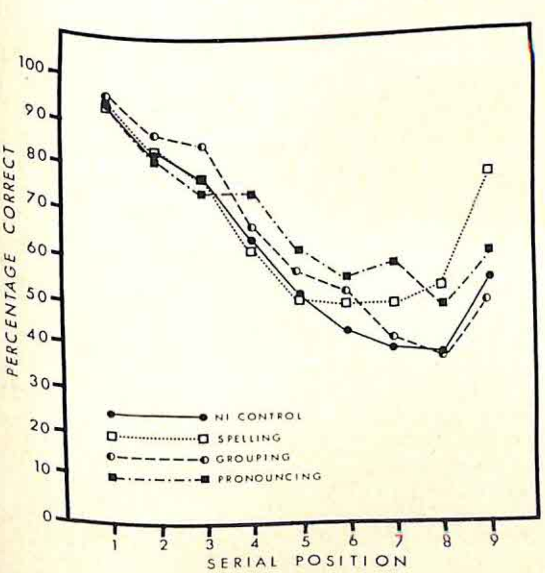


FIG. 2. Percentage of correct responses as a function of serial position and encoding instructions.

*Error data.*—Errors were classified into three mutually exclusive categories: omissions (no letter entered), misplacements (letter entered in incorrect serial position), and importations (entry of a letter not contained in the string). As might be expected from the correct score analysis, significantly more errors of each type occurred for the Low strings, $F$s (1, 72) $\geq 17.92$, $p < .001$. Separate analyses of variance for each error type failed to reveal significant effects of either encoding instructions or block order, although a significant Block Order × Encoding Instructions interaction was found for misplacement errors, $F$ (3, 72) = 3.67, $p < .05$, indicating fewer misplacement errors for the Low–High block order for groups using overt vocalization (Pronouncing and Spelling), while the two remaining encoding conditions exhibited slightly more misplacement errors for the Low–High block order.

## DISCUSSION

The present results replicate the digram-frequency effect found by Mayzner and Schoenberg (1964), viz., that immediate serial recall of high-frequency consonant strings is significantly better when the letters are arranged into digrams of high frequency than when the same letters are arranged into digrams of low frequency. The mechanisms underlying this effect, however, remain unclear. The failure to find any significant interaction between encoding condition and digram frequency (i.e., the failure of the encoding instructions to change the magnitude of the digram-frequency effect) made it unlikely that this effect could be the result of *S*s' adoption of an encoding strategy which is more efficient for strings of high than for strings of low digram frequency. Furthermore, differential effects of encoding upon High and Low strings were not found for the two ancillary variables: shape of the serial position curve and errors; this also argues against an encoding explanation of the digram effect.

It might be argued that the digram-frequency effect is the result of some kind of encoding operation other than those manipulated in the present study, i.e., that uninstructed *S*s spontaneously adopt some other encoding strategy, and this hypothetical other strategy is more efficient for the High strings. If this were the case, however, there should have been a significant Encoding Condition × Digram Frequency interaction, with the digram-frequency effect being greater for the NI control group than for the three groups instructed to use a different strategy. In short, nothing in the present data supports the hypothesis that the superior retention of high-frequency digram arrangements is mediated by a more efficient encoding operation.

Other purposes of this experiment were to determine whether the various encoding operations would have any facilitating effect on immediate serial recall of consonant strings and to evaluate the effect of presenting the strings in the High–Low versus the Low–High block order. The relevant findings may be summarized as follows: (*a*) Instructions to pronounce and instructions to spell aloud both facilitated performance, on both types of strings, when the Low–High block order was used but not when the reverse order was used, and (*b*) the High–Low block order facilitated performance, on both types of strings, for the NI control *S*s but not for the three groups instructed to encode. Convincing explanations for this set of results are not immediately apparent. These findings suggest, however, that two sources of facilitation were involved, one being overt vocalization and the other an unknown factor associated with the High–Low block order, and that their effects were not additive; i.e., facilitation from one of these two sources precluded any significant amount of facilitation from the other.

The finding that, at least in the Low–High block order, recall was improved by instructions to spell aloud confirms the results found by Murray (1966, 1967) who compared spelling aloud with silent reading of strings of eight consonants. The present study, in addition, shows improved recall as a result of overt vocalization regardless of whether it takes the form of spelling the string letter by letter or pronouncing the string as a collection of syllables. The serial position curves in the Murray studies are also similar to those in the present experiment in showing the improved recall which occurs as a result of overt vocalization to be localized largely in the latter half of the curve, especially in the terminal position.

While it is tempting to interpret the serial position data in terms of a two-phase model of memory (e.g., Atkinson & Shiffrin, 1968; Broadbent, 1958; Waugh & Norman, 1965), the limited recency effect which occurred only for the last item rather than for the last few items argues against a strong short-term memory component. Until immediate retention is compared with delayed retention, any assignment of the digram-frequency effect to either primary or secondary memory would be premature. Research currently in progress aims at pinpointing the elusive locus of the digram-frequency effect.

## REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.). *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Broadbent, D. E. *Perception and communication.* New York: Pergamon, 1958.

Laughery, K. R., & Pinkus, A. L. Recoding and presentation rate in short-term memory. *Journal of Experimental Psychology*, 1968, **76**, 636–641.

Lindley, R. H. Words and pronunciation as coding aids. *Psychonomic Science*, 1966, **6**, 395–396.

Mayzner, M. S., & Schoenberg, L. Single-letter and digram frequency effects in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 397–400.

Murray, D. J. Vocalization-at-presentation and immediate recall, with varying recall methods. *The Quarterly Journal of Experimental Psychology*, 1966, **18**, 9–18.

Murray, D. J. Overt versus convert rehearsal in short-term memory. *Psychonomic Science*, 1967, **7**, 363–364.

Palmer, S. E., & Ornstein, P. A. Role of rehearsal strategy in serial probed recall. *Journal of Experimental Psychology*, 1971, **88**, 60–66.

Ryan, J. Temporal grouping, rehearsal, and short-term memory. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 148–155.

Stark, K., & Calfee, R. C. Recoding strategies in short-term memory. *Journal of Experimental Psychology*, 1970, **85**, 36–39.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Wickelgren, W. A. Rehearsal groupings and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 97–102.

(Received August 6, 1971)

Journal of Experimental Psychology
1972, Vol. 93, No. 1, 111–117

# ACQUIRED PLEASANTNESS AND PAIRED-ASSOCIATE LEARNING IN MIXED AND HOMOGENEOUS LISTS[1]

ALBERT SILVERSTEIN[2]

*University of Rhode Island*

Four experiments are reported in which nonsense syllables first acquired differing degrees of pleasantness (PL) by being paired with either pleasant (P) or indifferent (I) pictures and were then learned as responses in paired-associate lists that were either homogeneous or mixed in PL. Performance with P-paired syllables was superior to that with I-paired syllables in homogeneous lists *only* when affective contrast was used during the syllable-picture pairing. But with mixed lists, performance with P-paired syllables was superior both with and without prior affective contrast. The advantage of the P-paired syllables in mixed lists was found mainly in response learning, while that advantage in homogeneous lists was found mainly in association learning and was of a lesser magnitude.

A large number of experiments have demonstrated that when neutral stimuli are paired with positively affective stimuli, the neutral stimuli will increase in judged pleasantness (PL) or positive evaluative meaning (e.g., Razran, 1938a, 1938b, 1940a, 1940b, 1956; Staats & Staats, 1957; Staats, Staats, & Biggs, 1958; Yavuz, 1963). The first demonstration that such pairing can also enhance performance in a learning task was made in experiments by Silverstein (1966) and Silverstein and McCreary (1964), who found a heightened level of performance in a paired-associate (PA) learning task for nonsense syllables that had previously been paired with pleasant (P) photographs, as compared with syllables that had previously been paired with indifferent (I) photographs. They also found that the P-paired syllables were judged to be more pleasant than the I-paired syllables.[3]

In the experiments of Silverstein (1966) and Silverstein and McCreary (1964), all *S*s learned a first list that contained both P and I pictures and a second list that contained both P-paired and I-paired syllables. In order to separate the mechanism by which the transfer of PL took place from the mechanism by which acquired PL operated on the learning of the syllables, it is necessary to discover whether a mixed list is necessary in the first and/or second task in order to obtain the superiority in speed of learning the P-paired syllables. The experiments reported in this paper pursue the mixed- versus homogeneous-list issue, while keeping the other conditions of research as comparable as possible to the previously cited experiments. In Exp. I and II, both the first and second lists were homogeneous with regard to PL. In Exp. III, affective contrast was isolated in the first list by having *S*s learn a first list that was mixed with regard to picture PL and a second list in which only the P-paired *or* the I-paired syllables were learned. In Exp. IV, affective selectivity was isolated in the second list by first giving *S*s a list in which either P or I pictures were learned as responses to the syllables and then presenting these syllables along with an equal number of nonpaired syllables in the second PA list.

---

[1] The research reported in this paper was supported by Grant MH 08974 from the National Institute of Mental Health. The author is indebted to Richard Dienstbier, Lionel Blatchley, and Alice Marshall for having run *S*s.

[2] Requests for reprints should be sent to Albert Silverstein, Department of Psychology, University of Rhode Island, Kingston, Rhode Island 02881.

[3] Unpublished data from our Laboratory indicate that the typical effect of acquired PL upon PA learning is not found with female *S*s unless special measures are taken to remove task anxiety. This led to the exclusive use of male *S*s in our other acquired-PL experiments while the nature of this difference was being investigated. These results will be reported in a subsequent paper.

## Experiments I and II

In both experiments, experimental *S*s first learned a PA list in which all the response terms were either P or I pictures and then learned a second PA list in which the syllables used as the prior stimulus terms became the responses for two-digit numbers. In Exp. I, both lists contained eight pairs in order to replicate all other details of a previous mixed-list experiment that had used the same *S* and photograph pools (Silverstein, 1966). Experiment II used only six pairs in each of the lists learned in order to replicate the second-list conditions of Exp. III and to make certain that any differences found in the results of Exp. I and III were not merely the consequence of the latter using shorter lists. Exposure-control groups in both experiments first rated the syllables' meaningfulness for a number of trials equal to the experimental groups' PA task before learning the syllables as responses to the numbers. These control *S*s established a base line of performance against which to evaluate the performance of the groups learning the I-paired syllables. The latter groups, in turn, indicated the level of performance that results from previously having learned the responses of a list as stimuli for terms of minimal PL value and established a base line against which to compare the performance of the groups learning the P-paired syllables.

*Method.*—In all details except list length, the details of Exp. I and II were identical. A total of 48 male *S*s in Exp. I and 72 male *S*s in Exp. II were recruited from introductory psychology classes and randomly assigned to one of three Task 1 conditions: pairing of syllables with pleasant pictures (EP), pairing of syllables with indifferent pictures (EI), and familiarization with syllables by successive rating trials (C). The *S*'s task in the two experimental conditions (EP and EI) was to learn to respond with his own short label for the correct photograph when he saw each syllable. Learning was carried to a criterion of two successive perfect trials.[4] Control *S*s were told to rate each syllable for its meaningfulness when it appeared. Each control *S* was given the same number of rating trials as the prior *S* in the EI condition. The second task for all *S*s was learning the List 1 syllables as responses to two-digit numbers to a criterion of a perfect trial or for 10 trials, whichever took longer. Finally, each *S* was instructed to rate the syllables for PL.

The P and I photographs were selected from a pool of 105 whose PL had been rated by an independent group of 85 male *S*s from the same population on a 9-point thermometer-type scale. The ranges for the ratings of the P and I pictures were 6.82 to 7.65 and 4.80 to 5.10, respectively. The syllables came from the 24% to 53% Archer (1960) values for meaningfulness, while the numbers were among the lowest meaningful in Battig and Spera's (1962) list. Both syllables and numbers were of low intralist similarity. All stimuli were presented by Kodak Carousel automatic slide projector for 5 sec. in both PA lists, with a 10-sec. intertrial interval. For control *S*s, a blank slide appeared for 5 sec. following each syllable in the rating task. In each PA list, an equal number of *S*s were given each of the four randomly determined orders of presentation on the first trial of each list. Experiment II was begun following the conclusion of Exp. I.

*Results and discussion.*—Table 1, below, shows the means of major results. It can be seen that there were virtually no differences between the EP and EI groups in trials to a perfect recitation for either List 1 or List 2 in both experiments. Nor were there any differences between these two groups in either the response-learning stage (trials to first response, correct or not) or the association-learning stage (trials from first response to first correct response; see Underwood & Schulz, 1960). On the other hand, the control group appears to have done slightly worse in Task 2 for both experiments. An analysis of variance was performed on the data of both experiments for number of correct responses in 10 trials on the second list. The value of $F$ (1, 114) associated with list length was 3.95, which is significant at the .05 level, but must be interpreted cautiously since the two experiments were not run concurrently. The $F$ for Task 1 conditions (2, 114) was 2.08, which is not significant, nor was there any significant interaction between Task 1 and list length. However, that portion of the Task 1 mean square associated with the experimental versus control group difference was significant: $F\ (1, 114) = 3.98,\ p < .05$. For the data

---

[4] Half of the *S*s in each condition of Exp. II were run to a criterion of one perfect recitation, to make their training equal to that of *S*s in Exp. III. Since no differences in List 2 performance were found between these subgroups, their data were pooled.

TABLE 1

MEANS OF MAJOR STATISTICS IN EXPERIMENTS I AND II

| Statistic | Exp. I | | | Exp. II | | |
|---|---|---|---|---|---|---|
| | EP | EI | C | EP | EI | C |
| List 1: trials to criterion | 10.50 | 10.56 | — | 7.67 | 7.71 | — |
| List 2: trials to criterion | 9.56 | 10.50 | 12.38 | 7.06 | 8.00 | 9.72 |
| Trials: response learning | 4.11 | 4.03 | 4.19 | 3.10 | 3.29 | 3.44 |
| Trials: association learning | .92 | .72 | 1.02 | .43 | .59 | .82 |
| PL ratings | 4.59 | 4.45 | 4.09 | 4.62 | 4.22 | 4.20 |

Note.—Experimental conditions are as follows: EP = pairing of syllables with pleasant pictures; EI = pairing of syllables with indifferent pictures; and C = familiarization with syllables by successive rating trials.

of Exp. I, a separate analysis was performed on just those syllables that had been paired with the four pictures in each list which were used previously in a mixed-list experiment (Silverstein, 1966) and no superiority was found for the P-paired over the I-paired syllables in trials to criterion.

There was no substantial difference between the EP and EI groups' PL ratings of the syllables (on a 7-point scale) following List 2 learning in either experiment, but in Exp. I both experimental groups seemed to rate the syllables higher on PL than did the control group. An analysis of variance for Exp. I showed a significant between-groups effect: $F\,(2, 45) = 4.15$, $p < .02$. Orthogonal analyses on the ratings showed that only the difference between the pooled experimental groups and the control group approached significance, $F\,(1, 45) = 3.71$, $p < .10$. For Exp. II, the between-groups effect fell between the .05 and .10 significance levels, $F\,(2, 69) = 2.78$, as did the difference between the EP and EI groups, $F\,(1, 46) = 3.79$. Since the PL ratings obtained from *S*s in these experiments were contaminated by prior learning of the syllables in List 2, two additional groups of 12 *S*s were run in the EP and EI conditions for List 1 and then asked to rate the syllables' PL. Once again, the difference failed to reach significance, EP = 4.84, EI = 4.41, $t\,(23) = 1.20$, $p > .20$.

These experiments demonstrated that the removal of affective contrast from both the affective-conditioning list and the performance-test list will eliminate the typical superiority of P-paired over I-paired syllables in both learning and rated PL. Since the typical superiority of first-list learning for P over I pictures (whose affective values are established preexperimentally) was also absent in these experiments, it seemed probable that affective contrast during performance is a necessary condition for the superiority of learning P responses. This hypothesis was tested in Exp. III, in which *S*s learned a first list that was mixed with regard to picture PL and a second list in which only the P-paired or the I-paired syllables were used.

## EXPERIMENT III

*Method.*—Fifty-six male *S*s were recruited from introductory psychology classes and randomly assigned either to a P or an I condition. In the first list, all *S*s learned 12 pairs of syllables and pictures to a criterion of one perfect recitation, half the pictures being those used in the EP condition of Exp. II and half being those used in the EI condition. This longer list was necessary because only half of its syllables were to be used for the second list, and the less stringent criterion was necessary because of the longer time needed to learn the longer list. In List 2, each *S* learned either the six P-paired or the six I-paired syllables as responses to two-digit numbers to a criterion of a perfect trial or for six trials, whichever took longer. One *S* in each group failed to reach criterion within 16 trials and was replaced. Within both groups, an equal number of *S*s was randomly assigned to each of four different syllable-picture pairings, such that each syllable was equally often P- and I-paired; *S*s were successively assigned to each of four random orders of presentation for the first trial of both lists. In all other details, the method was identical to that of Exp. II.

*Results and discussion.*—That the two groups of *S*s were of equal learning ability was shown by the nearly identical means in trials to criterion for List 1, P = 9.15 versus

I = 9.81, $t(54) < 1$. Both groups learned the P pictures significantly faster than the I pictures, the superiorities in percentage-correct responses being 8.8% for Group P, $t(27) = 3.76$, $p < .01$, and 6.9% for Group I, $t(27) = 3.09$, $p < .01$.

While the performance of the P group in List 2 was somewhat superior to that of the I group, as measured by trials to criterion (6.52 vs. 8.11), this difference was only marginally significant, $t(54) = 1.81$, $.05 < p < .10$. Closer analyses revealed that there was no difference between the two groups in trials to complete response learning (P = 3.46 vs. I = 3.51, $t < 1$) but that Group P showed significant superiorities in both trials to complete association learning (P = .19 vs. I = .70 trials, $t = 2.04$, $p < .05$) and percentage correct following first correct response (P = 91.1 vs. I = 83.2, $t = 2.55$, $p < .02$). In addition, Group P showed both a significantly smaller *rate* of intralist errors than Group I (31.1% vs. 45.9%, $t = 2.14$, $p < .05$) and a significantly higher PL rating for the syllables following the conclusion of learning (4.97 vs. 4.11, $t = 2.44$, $p < .02$).

Contrary to the hypothesis, the results of this experiment point strongly to the conclusion that, provided affective contrast is present in the affective-conditioning list, P-paired syllables will be learned more rapidly and judged more PL than I-paired syllables even though affective contrast is absent in the performance-test list. But it should be noted that the pattern of results obtained in Exp. III is unlike that obtained in studies using mixed lists for both tasks (Silverstein, 1966). In the present experiment, the superiority of the P-paired syllables was smaller than that found in mixed second lists, and its locus was shifted from being predominantly based upon response learning to later in the learning sequence, i.e., to faster and firmer association learning. Moreover, *S*s in this experiment showed greater confusion of the I- than the P-paired syllables, a tendency not found in mixed-second-list experiments.

## EXPERIMENT IV

The preceding experiments have shown that while the acquired-affect phenomenon is eliminated by making both the first and second lists homogeneous, a small effect of acquired PL can be obtained when only List 1 contains both P and I items for each *S*. The next step was to determine whether a first list which is homogeneous with regard to PL and a second list which is heterogeneous will produce a superiority of P-paired over I-paired syllables. In two concurrent experiments, IV-A and IV-B, *S*s learned a first list of six pictures, either homogeneously P or homogeneously I, as responses to syllables and were then given these syllables along with six nonpaired syllables in the second PA list as responses to the two-digit numbers. In Exp. IV-A, the nonpaired syllables were new to *S*, while in Exp. IV-B *S* received familiarization trials with the nonpaired syllables between the learning of the two lists.

*Method.*—Sixty-four male *S*s were recruited from introductory psychology classes and assigned randomly to the P and I conditions of either Exp. IV-A or IV-B. In List 1, *S*s received either the six P or the six I pictures used in Exp. II as responses to the syllables. Each *S* received a randomly determined half of the 12 syllables used in Exp. III for List 1 and the other half as the nonpaired syllables. For each set of syllables used, there were two different pairings with the pictures, such that each syllable was paired equally often with a P or I picture. In the second list, there were two

TABLE 2

MEANS OF MAJOR STATISTICS IN EXPERIMENTS IV-A AND IV-B

| Statistic | P Group | | I Group | |
|---|---|---|---|---|
| | Paired syllables | Nonpaired syllables | Paired syllables | Nonpaired syllables |
| Exp. IV-A | | | | |
| List 1: No. correct in 8 trials | 29.13 | — | 24.25 | — |
| List 2: No. correct in 12 trials | 35.88 | 31.69 | 25.50 | 30.56 |
| Trials: response learning | 4.54 | 5.20 | 5.76 | 5.18 |
| Trials: association learning | .61 | 1.02 | .64 | .69 |
| Response recall: no. correct | 3.88 | 3.13 | 2.50 | 3.00 |
| PL ratings | 4.39 | 3.08 | 4.02 | 3.81 |
| Exp. IV-B | | | | |
| List 1: no. correct in 8 trials | 28.50 | — | 23.88 | — |
| List 2: no. correct in 12 trials | 30.56 | 33.31 | 27.88 | 34.63 |
| Trials: response learning | 4.64 | 4.28 | 5.42 | 4.67 |
| Trials: association learning | 1.11 | 1.19 | 1.06 | .64 |
| Response recall: no. correct | 3.13 | 3.63 | 2.75 | 3.88 |
| PL ratings | 3.94 | 4.14 | 3.98 | 4.02 |

different syllable-number pairings, such that each number was paired equally often with a paired or nonpaired syllable.

All *S*s were run for 8 trials on List 1 and 12 trials on List 2. The *S*s in Exp. IV-A went directly from List 1 to List 2 learning, while *S*s in Exp. IV-B were given four familiarization trials between the two lists, in which each syllable appeared for 5 sec., and *S* noted its "difficulty of learning" on a separate page of a special booklet. The total time for this task was 5 min. Four different random orders of presentation were used for both List 1 and the familiarization task and three random orders were used for List 2, with *S*s assigned successively to each order in each treatment condition. In all other details the method was identical to that of Exp. I.

*Results.*—Table 2 shows the major results of both experiments. The number of correct responses in List 2 was higher for the P-paired than for the new syllables but lower for the I-paired than the new syllables in Exp. IV-A. When the nonpaired syllables had been familiarized (Exp. IV-B), performance with them improved and performance with the paired syllables deteriorated in both the P and I conditions, which is not surprising since the familiarization trials occurred just prior to List 2 learning. But the results of Exp. IV-B are congruent with those of IV-A, since the superiority of the nonpaired syllables was much smaller in the P than in the I condition. An overall analysis of variance revealed no significant effects, however, and only the interaction between paired versus nonpaired syllables and PL of the pairing came close, $F\ (1, 60) = 2.56$, $p < .15$. A separate analysis performed on Exp. IV-A produced a significant interaction of PL and pairing versus nonpairing: $F\ (1, 30) = 6.55$, $p < .02$. Correlated-means $t$ tests on the paired versus nonpaired differences produced marginally reliable values in both the P group, $(15) = 1.80$, $p < .10$, and the I group, $(15) = 2.05$, $p < .10$.

The preceding pattern was repeated reliably in the response-learning stage. While an overall analysis of variance showed no significant effects, a separate analysis of Exp. IV-A again showed a significant interaction of PL and pairing versus nonpairing, $F\ (1, 30) = 4.19$, $p < .05$. It was not possible to analyze the association-learning stage for all *S*s, since only 42 of them reached a perfect recitation by the end of training; however, analyses performed on the data of the fastest 8 *S*s per condition showed no significant effects. To remedy this lacuna, each *S* was given either a response-recall test or an associative-matching test between Trials 6 and 7 of anticipation learning, unreinforced in both cases.[5] The associative-matching test produced no significant or substantial differences among any of the experimental treatments. On the other hand, the response-recall test produced a significant interaction of PL with pairing versus nonpairing in Exp. IV-A, $F\ (1, 14) = 6.10$, $p < .05$, and a significant superiority for the nonconditioned (familiarized) items in Exp. IV-B, $F\ (1, 14) = 4.89$, $p < .05$ (see Table 2). No significant differences were found in the rate of misplaced responses in the various conditions.

The PL ratings for the syllables following List 2 learning produced only one significant difference, that between P-paired and nonpaired syllables in Exp. IV-A, $t\ (15) = 3.01$, $p < .01$.

These data make it clear that affective contrast during the acquisition of PL by nonsense syllables is not a necessary condition for those syllables to show a subsequent superiority over neutral syllables *in a mixed list*. This was particularly clear in Exp. IV-A, where the beneficial effects of acquired PL were not pitted against recent familiarization. Postexperimental interviews revealed that *S*s generally adopted a strategy of attempting to learn first those syllables that had not been part of a prior learning task (Battig, Allen, & Jensen, 1965), a strategy which was aided by the familiarization trials given in Exp. IV-B.

## GENERAL DISCUSSION

Two empirical generalizations can be drawn from the preceding experiments. The first is that affective contrast in either List 1 or List 2 is a sufficient condition for P-paired syllables to be learned more rapidly than I-paired syllables. Underwood (1966, pp. 485–486) has argued cogently that a variable which is effective only in mixed-list designs operates by engaging *S*s' momentary selection strategies, while a variable that truly alters the difficulty of the list will also operate when different values of that variable are assigned to separate, homogeneous lists. Following this argument, it would be concluded that the acquired-PL effect obtainable only after affective contrast in the syllable-picture pairings is based upon the P-paired items being truly easier to learn than the I-paired items. On the other hand, that acquired-PL effect which can be obtained with or without prior affective contrast is one that is based upon *S*s' selective-performance dispositions ("priority habits").

---

[5] The *S*s given these two different tests performed comparably on all other measures, and their data were pooled.

The foregoing analysis strongly suggests that there are two distinct mechanisms by which acquired PL can operate on performance, one producing a disposition for *S*s to select P-paired over I-paired items for earlier learning and the other producing a genuine, superior learnability for the P-paired items. Such a duality is strongly supported by the second empirical generalization to be drawn from these experiments: that when an acquired PL effect was obtained in a homogeneous test list it was a small one whose locus was exclusively within the association-learning stage, as contrasted with those effects obtained in mixed test lists which were substantially larger and were located primarily in the response-learning stage. Ekstrand (1966) has cautioned that the stage analyses developed by Underwood & Schulz (1960) are sensitive to *S*'s willingness to guess when he is uncertain of a response and, therefore, may misrepresent the stages of learning. Thus, any variable that can affect *S*'s confidence (e.g., intralist similarity) will artificially alter the size of the response-learning stage and, therefore, the association stage of a PA task measured in this way. However, it seems unlikely that acquired PL would influence *S*'s confidence in guessing and evidence is available from unpublished experiments that P-paired syllables are not judged to be easier to learn than I-paired syllables. Moreover, our *S*s were all instructed to guess whenever they were not sure of a response and subsequent interviews indicated that they had followed these instructions. Finally, it was shown in Exp. IV that differences found in a stage analysis were paralleled by differences between response-recall and associative-matching tests, tests which were advocated by Ekstrand to measure the stages of learning.

The experiments reported here do not allow a definitive description of the two aforementioned mechanisms. But it is reasonable to hypothesize that the disposition to select syllables which have acquired P characteristics for faster response learning than I syllables is based upon an enhanced rate of implicit rehearsal of representational responses to them. Since it is rare that *S*s will attempt to distribute their efforts evenly across all the items on a PA list from the beginning of training, those items which *S* rehearses more vigorously would come through the response-learning stage more rapidly, but would not necessarily be associated with their appropriate stimuli more rapidly. On the other hand, the faster association learning and reduction in error tendencies for P-paired syllables found in homogeneous test lists is likely to be based upon their having acquired a more reliable capacity to evoke images that are relevant to the pictures they were paired with than did the I-paired syllables. This would allow such images to serve as more distinctive and effective mediators in the formation of subsequent associations (Paivio, 1969) in the P than in the I case. This conclusion is supported by unpublished data on the nature of the associations given to P- and I-paired syllables prior to their being learned in a second list. The formation of such distinctive associative mediators for the P-paired syllables seems to require some form of discrimination training, since the association-learning superiority of the P-paired syllables in homogeneous lists was not found without prior affective contrast in List 1 (Exp. I and II).

The PL ratings of the syllables showed that P-paired items were judged more pleasant than neutral items if either the first or second list was mixed, indicating that the *transfer* of PL to neutral items does not require affective contrast so long as contrast is available in the test of such transfer. This may be interpreted according to Peters' (1935) "judgmental theory of feeling," which emphasized that affective ratings were judgments of the degree to which one would normally approach a stimulus. Such judgments were said to require *S*'s discrimination of a stimulus' approach value from that of some normative value, and removing affective contrast from both lists provides *S* with no basis for making such discriminative judgments.

## REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**(10, Whole No. 497).

Battig, W. F., Allen, M., & Jensen, A. R. Priority of free recall of newly learned items. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 175–179.

Battig, W. F., & Spera, A. Rated association values of numbers from 0–100. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 200–202.

Ekstrand, B. A note on measuring response learning during paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 344–347.

Paivio, A. Mental Imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

Peters, H. N. The judgmental theory of pleasantness and unpleasantness. *Psychological Review*, 1935, **42**, 354–386.

Razran, G. H. S. Music, art, and the conditioned response. *Psychological Bulletin*, 1938, **35**, 532–541. (a)

Razran, G. H. S. Conditioning away social bias by the lunch-room technique. *Psychological Bulletin*, 1938, **35**, 693–699. (b)

Razran, G. H. S. Determinants of the consolidation or *Pragnanz* of conditioned preferences. *Psychological Bulletin*, 1940, **37**, 481–492. (a)

Razran, G. H. S. Conditioned response changes in rating and appraising socio-political slogans. *Psychological Bulletin*, 1940, **37**, 564–571. (b)

Razran, G. H. S. The conditioned evocation of attitudes. *Journal of Experimental Psychology*, 1956, **48**, 274–282.

Silverstein, A. Acquired word value, verbal learning, verbal learning and retention: Replication and extension. *Psychological Reports*, 1966, **18**, 357–358.

Silverstein, A., & McCreary, C. Acquired word value and verbal learning. *Psychological Reports*, 1964, **14**, 499–504.

Staats, A. W., Staats, C. K., & Biggs, D. A. Meaning of verbal stimuli changed by conditioning. *American Journal of Psychology*, 1958, **71**, 429–431.

Staats, C. K., & Staats, A. W. Meaning established by classical conditioning. *Journal of Experimental Psychology*, 1957, **54**, 74–80.

Underwood, B. J. *Experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts, 1966.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. New York: Lippincott, 1960.

Yavuz, H. S. The retention of incidentally learned connotative responses. *Journal of Psychology*, 1963, **55**, 409–418.

(Received August 7, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 118–122

# EFFECT OF PARTIAL RECALL ON THE RANSCHBURG PHENOMENON[1]

CONNIE JO HARRIS[2] AND JOHN C. JAHNKE

*Miami University*

Stimuli were strings of eight digits that contained either all-different digits or a digit in Serial Position (SP) 2 that was repeated in either SP 5 or 6. Immediate retention was measured either by a partial-recall technique that required *S* to omit only the digit in SP 2, or by the usual method of complete recall. Recall was poorer for repeated than corresponding control digits with complete recall (the Ranschburg effect, RE); however, repeated and control digits were equally well recalled under conditions of partial recall. These results are consistent with the view that conditions reducing output interference also reduce the inhibitory effects of intraserial repetition.

One aspect of the Ranschburg effect (RE) lies in the omission or incorrect reproduction of a repeated element in a string of otherwise all-different elements presented for immediate recall (cf. McGeoch, 1942). Interpretations of this effect have drawn attention to the influence of input and output processes. Among these interpretations, those of Obonai and Tatsuno (1954) and Wickelgren (1965, 1966) emphasize the role of input (attentional, perceptual, and encoding) processes and account for the RE in terms of *S*'s failure to detect or to encode appropriately the repeated item. First, it has been held that *S* may fail to detect the occurrence of a repetition within a string and may, therefore, be biased against including a repetition in recall (cf. Obonai & Tatsuno, 1954; Wickelgren, 1965, 1966). Second, it has been held that *S* may fail to recode strings in terms of their repetition structure, say, a "run of three 2's" (Wickelgren, 1965) and may, therefore, be less likely to emit the repeated items in recall as often as he should.

More recent work indicates that output (memory and retrieval) processes may also be substantial contributors to the RE. Crowder (1968) found that when *S* was required to include a prearranged consonant prefix in his recall, the probability of recall of that letter was reduced when it also occurred in the to-be-remembered series. Since a repeated letter was not included in the stimulus series during presentation but was supposed to be included during recall, Crowder's results were taken to indicate that the inhibitory effects associated with intraserial repetition of elements can occur during recall.

The usual result in studies of the RE is that more errors are associated with the second occurrence of the repeated element. It is possible that interference effects associated with the delayed output of that item contribute to this result. Consistent with this view are data showing a shift of the focus of the inhibitory effect from the second to the first occurrence of the repeated element under conditions of reversed recall (Items 5–8 in order, followed by 1–4 in order) of eight-letter strings (Jahnke, 1969a). Also consistent are data from studies of probed recall (Jahnke, 1970) and recognition memory (Wolf & Jahnke, 1968) that show performance on repeated elements at least as good as on control elements. If output interference (cf. Tulving & Arbuckle, 1966) were an important contributor to the RE, as these data suggest, any technique that reduces output interference should reduce the RE.

The present study was undertaken to test further the possibility that the Ranschburg phenomenon may be, at least in part, a function of the retrieval procedure. Although digit series containing a repeated element were presented, *S* was required not to recall the first occurrence of that repeated element in one of the experimental conditions. It was considered that such conditions would reduce output interference, and, thereby, lessen the magnitude of the RE.

---

[1] Based on a thesis presented by the first author to the faculty of Miami University in partial fulfillment of the requirements for the MA degree. This work was supported in part by Grant GB-8106 from the National Science Foundation to Miami University.

[2] Requests for reprints should be sent to Connie Jo Harris, Department of Psychology, Miami University, Oxford, Ohio 45056.

## METHOD

*Subjects.*—The *Ss* were 24 introductory psychology students attending Miami University. They participated in the experiment to fulfill a course requirement.

*Materials and procedure.*—Basic stimuli were strings of eight digits derived from a table of random numbers with the restriction that no digit was repeated in any string. Experimental strings were derived from these by substituting the appropriate (repeated) digit in the appropriate serial position. All experimental strings included a single repetition of a single digit.

The stimulus series were randomly ordered into two separate blocks of 28 strings (trials) each. Half of the strings in each block were control strings (no repeated items), and half were experimental strings (a single repeated item). Half of the experimental strings included a repeated element in Serial Positions (SPs) 2 and 5, and the other half, SPs 2 and 6. The order of occurrence of experimental (E) and control (C) strings in each block was counterbalanced in the *abba* order. The order of experimental strings was randomized within each block of four trials. The 28 strings of one block were recalled after instructions that emphasized that *S* recite out loud the digits in their original order immediately after he had heard each series (Complete-Recall condition). The 28 strings of the other block were recalled after instructions that were similar, except that they emphasized that *S* was always to omit the digit in SP 2 from his recall (Partial-Recall condition). All *Ss* received both blocks of 28 trials. Half of the *Ss* received the stimulus blocks in the order complete followed by partial recall (Cond. C-1 and P-2, respectively); the other half, partial followed by complete recall (Cond. P-1 and C-2, respectively). The first four strings of each block were practice trials, and data from these were discarded.

The stimuli were presented aurally by a Wollensak Model 1500 tape recorder at the rate of 2/sec. The *E* recorded *S*'s oral recall, and the intertrial interval was approximately 10 sec.

## RESULTS AND DISCUSSION

*Overall recall.*—It will be remembered that in partial recall, *S* was to recall items only in SP 1 and SPs 3–8; accordingly, only the corresponding items were included in the scoring of overall recall for Complete-Recall conditions. The proportions of completely correct recall (all items of a string recalled in their proper SPs) were somewhat lower for the combined experimental than control strings in the Complete-Recall condition (.406 and .455, respectively) and were slightly higher for experimental than control strings in the Partial-Recall condition (.396 and .368, respectively). However, a two-tailed direct-difference $t$ test of the difference between the Complete- and Partial-Recall conditions on this measure (the completely correct recall of experimental relative to control strings) only approached significance, $t\ (23) = 1.84$, $.05 < p < .10$.

*Performance on critical elements.*—The data of primary interest in the present study are those for the critical elements (repeated elements in experimental strings and elements occupying corresponding SPs in control strings). These data are presented separately for each recall condition and stage of the experiment in Fig. 1 in the form of serial-position analyses, for which items were scored correct even if misplaced. This scoring procedure avoids any artifactual advantage accruing to experimental strings with an ordered-recall scoring procedure, with which transpositions of repeated items cannot be detected but transpositions of control items can. In those instances in which a repeated digit was misplaced in recall, it was scored to the nearest SP in which a repeated element should have been. When only one occurrence of the repeated item was emitted in the recall of 2 & 6 strings and that occurrence was misplaced midway between SPs 2 and 6, half, at random, were scored to SP 2 and half to SP 6. There were very few such instances (a total of 4) under conditions of complete recall, but a larger number (a total of 18) under conditions of partial recall. With the condition of partial recall, the RE can be assessed only at SP 5 in the 2 & 5 condition and SP 6 in the 2 & 6 condition. Therefore, a larger number of repeated elements that were actually emitted in recall were counted as errors in partial recall than in complete

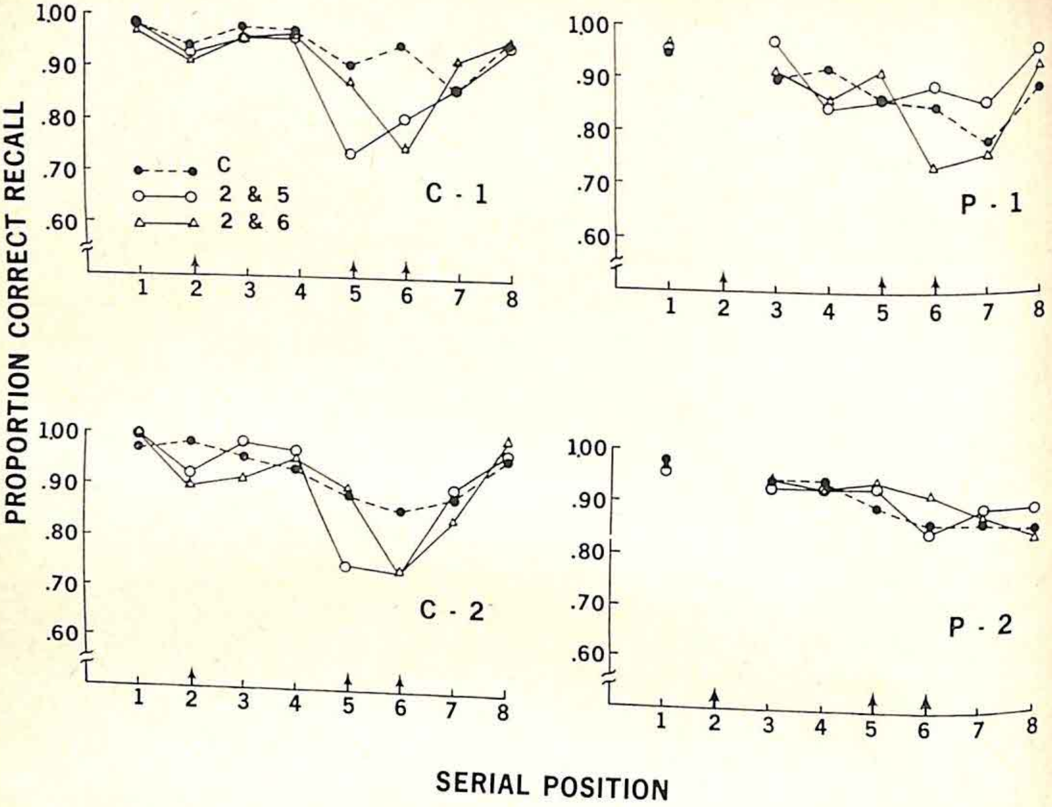


Fig. 1. Serial position analyses for Complete- and Partial-Recall conditions for each order of presentation.

recall. The effect of this procedure is, of course, to provide a conservative estimate of the facilitating effect of partial recall.

Turning now to Fig. 1, it can be seen that for Cond. C-1 and C-2, inhibition of recall of repeated items relative to control items (the RE) was found in both 2 & 5 and 2 & 6 strings. This result confirms the RE under the usual conditions of complete recall. The data of the Partial-recall conditions also generally conform to expectation. For Cond. P-1, performance on repeated items was the same as on corresponding control items in 2 & 5 strings (absence of the RE). For this condition, there was some RE for 2 & 6 strings, but, for the reason given immediately above, at least a part of this effect appears to be artifactual. Finally, for Cond. P-2, repeated items were recalled slightly, but nonsignificantly, better than corresponding control items in both 2 & 5 and 2 & 6 strings.

The differences between the recall of critical items in experimental and control strings were assessed for each recall condition separately by two-tailed Wilcoxon matched-pairs signed-ranks tests. For the obvious reason that *S*s in the Partial-Recall conditions were requested to omit the first of the critical items, data from only second occurrences were considered in these analyses. For Cond. C-1, a significant RE was found for 2 & 5, 2 & 6, and combined experimental strings, $T = 7.5$, $N = 11$, $p < .05$; $T = 5$, $N = 11$, $p < .01$; and $T = 2.5$, $N = 12$, $p < .01$, respectively. Results for Cond. C-2 were similar, except that the RE for 2 & 6 strings was of borderline significance, $T = 6$, $N = 11$, $p < .02$; $T = 16$, $N = 12$, $.05 < p < .10$; and $T = 7$, $N = 11$, $p < .02$, for the comparisons given in the same order. For Cond. P-1 and P-2, respectively, comparable results were, $T = 32$, $N = 11$, $p > .05$; $T = 12$, $N = 12$, $p < .05$; $T = 15.5$, $N = 11$, $p$

$> .05$; and $T = 23.5$, $N = 12$; $T = 20.5$, $N = 12$; $T = 18$, $N = 11$, all $p$s $> .05$. Thus, in only one of the six comparisons involving partial recall was there a significant RE (2 & 6 strings in Cond. P-1). However, it was mentioned above that at least part of the RE for 2 & 6 strings in Cond. P-1 appears to be a scoring artifact. It is possible to partial out this artifact by eliminating the data from those strings in which, when a single repeated item was recalled, the item was given in SP 4. When this was done, the RE was no longer significant, $T = 14.5$, $N = 11$, $p > .05$.

These data provide the major result of the present study: whereas the recall of repeated digits was poorer than that of corresponding control items under conditions of complete recall, repeated and control digits were equally well recalled under conditions of partial recall. While the empirical result seems clear, its interpretation is less so. It is true that the present findings are consistent with the notion that the inhibitory effects of intraserial repetition on recall are lessened under conditions that reduce the number of items to be output and thereby reduce output interference.

However, it might also be argued that the RE was attenuated under the Partial-Recall conditions simply because the task was easier when *S* had one less item to recall.[3] If this were the case, the overall correct recall of all items in SPs 1 and 3–8 in Cond. P-1 and P-2 should have been better than that of corresponding items in Cond. C-1 and C-2. To the contrary, as Fig. 1 suggests, overall performance was better under conditions of complete than partial recall, and the difference between Cond. C-1 and P-1 and between C-2 and P-2 did not approach significance, $t\ (22) = 1.43$ and $.14$, respectively, $p > .05$.

There is yet another interpretation that might be placed upon the present results. For example, the argument might be offered that conditions of partial recall predispose *S* to learn that some strings contain repeated elements and that the first occurrence of the repeated elements is always in SP 2, the item that is to be omitted in recall. Under these circumstances, *S* might be expected to set himself to listen for repetition during stimulus presentation and to be sure to include the repeated element in his recall. Thus, the RE would be lessened, if not eliminated, in this condition. This sort of interpretation focuses heavily upon the role of input processes and is, of course, opposed to the line of thought outlined in the first section of this paper. However, if *S*s who received partial recall in the first half of the session were to have adopted the processing strategy suggested by this "input" interpretation, performance on repeated items should have been improved in the second half of the session, which was given under instructions for complete recall. The data of Fig. 1 gave no evidence for such an effect. There was a significant RE in each of the Complete-Recall conditions, and their magnitudes were not significantly different, $t\ (23) = .79$, $p > .20$.

Finally, the argument could be made that *S*s selectively ignore the element in SP 2 during list presentation; therefore, experimental strings become functionally equivalent to control strings. Postexperimental questioning of each *S* disclosed that some *S*s apparently found it "hard to forget" the item in SP 2. On the other hand, about half of the *S*s reported attempting to "ignore" the item in SP 2 during presentation. Even these *S*s, however, reported that the attempt to ignore the item took "effort" or "time" or that they tried to "concentrate on" or "think about" leaving out that item. Such comments suggest that it was unlikely that the critical item was not registered in memory.

The present study is the logical inverse of the study by Crowder (1968), in which repetitions were not included in the stimulus series during presentation but were to be included by *S* in his recall. These conditions (no redundancy at input, redundancy at output) resulted in inhibition of the elements repeated in recall, whereas the Partial-Recall conditions of the present study (redundancy at input, no redundancy at output) resulted in no relative gain or loss of the elements repeated in presentation but not in recall. The considerations outlined here predispose us to the suggestion made initially that the inhibitory effects associated with intraserial repetition of elements are, at least in part, a result of interference arising from the act of recall. It must be recognized, however, that further research will be required to establish firmly this interpretation. Particularly desirable would be studies in which *S* was excused from recalling a nonrepeated item in a string that contained a repetition. Such a condition would show whether a presumed reduction in output interference independent of the repeated element itself would also reduce the RE. Finally, a boundary condition of the present study is the use of only two loci of repetition in experimental strings. Although it has been shown (Jahnke, 1969b) that an RE is obtained even when *S* knows that the string presented next will have a repeated item, a study incorporating a less predictable repetition structure would seem useful.

---

[3] We are grateful to the anonymous consulting editor who made this suggestion and whose comments have improved this report.

## REFERENCES

Crowder, R. G. Repetition effects in immediate memory when there are no repeated elements in the stimuli. *Journal of Experimental Psychology*, 1968, **78**, 605–609.

Jahnke, J. C. Output interference and the Ranschburg effect. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 614–621. (a)

Jahnke, J. C. The Ranschburg effect. *Psychological Review*, 1969, **76**, 592–605. (b)

Jahnke, J. C. Probed recall of strings that contain repeated elements. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 450–455.

McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, & Company, 1942.

Obonai, T., & Tatsuno, Ch. Facilitation and inhibition in memorizing a series of digits or letters. *Japanese Psychological Research*, 1954, **1**, 1–8.

Tulving, E., & Arbuckle, T. Y. Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, 1966, **72**, 145–150.

Wickelgren, W. A. Short-term memory for repeated and non-repeated items. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 14–25.

Wickelgren, W. A. Associative intrusions in short-term recall. *Journal of Experimental Psychology*, 1966, **72**, 853–858.

Wolf, T. M., & Jahnke, J. C. The effects of intraserial repetition on short-term recognition and recall. *Journal of Experimental Psychology*, 1968, **77**, 572–580.

(Received August 7, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 123–129

# CUED PARTIAL RECALL OF CATEGORIZED WORDS[1]

TIM DONG[2]

*Bucknell University*

The relationship between category recall and within-category recall was investigated in whole- and part-list recall conditions. On the basis of poststimulus cuing, *S*s were required to recall from all the categories or from a randomly selected single category of the categorized word list. Within-category recall was compared for whole- and part-list conditions; the results of the two experiments showed that partial recall produced greater within-category recall. Experiment II extended the generality of the partial recall effect in a situation controlling for nominal recall time. Experiment II further revealed that the partial recall effect (*a*) was limited to the first category cued and (*b*) occurred with categorically blocked input and not with random word input. This study points out the importance of retrieval factors in recall as well as suggesting a method which may be useful in studying within-category recall.

Several studies of recall of categorized word lists have shown the independence of recall of categories and recall of category instances. Cohen (1966) reported the invariance of the mean number of words recalled from a category over many variations of experimental conditions. For example, Cohen found that given a constant number of items per category at input, *S*s would remember more total words as list length increased from 5 to 10 to 15 categories. However, this increase in total words recalled reflected the increase in the number of categories recalled; the mean number of words per recalled category was invariant over the three list lengths. Tulving and Pearlstone (1966) substantiated Cohen's findings with different list lengths and numbers of items per category combinations. In addition, they also showed that mean words per category recalled was invariant over conditions of cued and noncued recall. Tulving and Pearlstone showed that total recall can be substantially increased by providing the list category names as cues during recall. The increase in words recalled in the cued condition was due to an increase in the number of categories recalled; the mean number of words per category recalled was not different between cued and noncued recall conditions.

Tulving (Tulving, 1968; Tulving & Pearlstone, 1966) has suggested that the invariance of within-category recall indicates two processes in the recall of organized information: (*a*) higher order units are recalled, and (*b*) information from within the higher order unit is recalled; these two processes are independent. Thus mean number of words per category recalled (within higher order unit recall) should not be influenced by such variables as list length which presumably affect category recall (higher order unit recall). Tulving's hypothesis is based on studies which have used recall of whole word lists.

The present study used a method of poststimulus cuing of partial recall to investigate within-category recall. The basic procedure of the partial recall method involves showing *S* a categorized word list with instructions to expect recall of the entire word list. At recall, however, *S* is cued to recall from a single randomly chosen category only. Thus, individual categories can be considered in recall without regard to the other categories of the list. The number of words per category recalled of the *S*s in the partial recall condition can then be compared with the mean number of words per category recalled of *S*s in the whole-list recall conditions.

---

[1] This article is based on a dissertation submitted to the University of California, Riverside, August 1968. The author wishes to thank the dissertation committee, Walter Kintsch, Chairman, Ramon J. Rhine, and Paul C. Rosenblatt, for their advice and guidance during the course of this research. This manuscript was prepared while the author was a United States Public Health Service Postdoctoral Fellow supported by the National Institute of Child Health and Human Development (Grant 1 FO2 HD-39235-01).

[2] Requests for reprints should be sent to Tim Dong, Department of Psychology, Bucknell University, Lewisburg, Pennsylvania 17837.

The specific purpose of the experiments presented here was to study the effects of part-list recall on the category and within-category recall relationship. If the recall of a category means that the contents of the category are maximally accessible, then there should be no difference in the mean number of words per category recalled regardless of part- or whole-list recall conditions. However, if partial recall is better than whole-list recall, this would indicate that within-category recall can be enhanced independently of category recall.

## EXPERIMENT I

### *Method*

*Subjects.*—The 45 *S*s were introductory psychology students at the University of California, Riverside, who served for class credit. The *S*s were randomly assigned to the experimental conditions; the assignments were such that the *n* for any group did not differ by more than 1 *S* from any other group at any point in time.

*Word list.*—The word list contained 100 words, each from 10 different word categories. The word categories were obtained from the Connecticut Norms (Cohen, Bousfield, & Whitmarsh, 1957); the categories chosen were judged by *E* to be nonoverlapping and to meet the criterion of having 10 words with taxonomic frequencies within the range of 9 to 100. The mean frequency of the words was 34.6 with a standard deviation of 21.3. The words were typed in capitals on 4 × 6 in. index cards.

*Procedure.*—The *S*s were tested in groups of no more than three persons. The *E* read instructions which explained the free recall method and the general experimental procedure to *S*s before word list presentation. The instructions did not reveal the categorized nature of the words, the number of words, or the existence of different experimental conditions. The *E* then manually presented the 100 words to *S*s at a practiced rate of 1 word per 2 sec. The *S*s saw three different random orders of the words in succession; the word orders were the same for all *S*s. Immediately after list presentation, *S*s had 30 sec. to read the recall instructions which were specific to each recall condition. Thus an interval of 30 sec. of instruction reading intervened between the end of stimulus presentation and the beginning of recall. The *S*s had 15 min. for written recall.

Three groups of *S*s were distinguished from each other on the basis of the type of poststimulus instructions given just before recall. The Noncued Recall group was given standard free recall instructions; the Cued Recall group was given a list of the category names and was instructed to use the names as aids in recall; the category names were always available during recall. The Partial Recall condition was instructed to recall only from a single specified category out of the 10 categories of the word list. The category from which any one Partial Recall *S* recalled was chosen randomly; each of the 10 categories were sampled at least once. Five categories were sampled twice in order to accommodate the 15 Partial Recall *S*s.

### *Results and Discussion*

The most important aspect of the data is the within-category results. The mean number of words per category recalled for the Noncued Recall group was 6.45; for the Cued Recall condition it was 6.70, and for the Cued Partial Recall condition it was 8.53. Mean number of words per category recalled was obtained by dividing the total number of words recalled by the number of different categories represented by the words recalled. In the Partial Recall condition, the mean number of words per category recalled was based on the single category each *S* had to recall. The difference between the means was statistically reliable, $F\ (2,\ 42) = 10.16$, $p < .001$. The superior recall of the Partial Recall condition shows that within-category recall is greater if *S* recalls from part of a word list instead of from the whole word list. This outcome is contrary to the hypothesis which suggests that within higher order unit recall is independent of higher order unit recall. Within-category recall can be increased beyond what is normally recalled in a whole-list recall situation by asking for partial recall.

The two whole-list recall conditions did not differ in total words recalled or in the number of categories recalled. The mean total words recalled was 63.47 and 67.00, while the mean number of categories represented by the recalled words was 9.80 and 10.00 for the Noncued Recall and the Cued Recall groups, respectively. The presence of category names as recall cues for the Cued Recall condition was ineffective in increasing recall beyond that of the Noncued Recall condition. The three presentations of the word list and the 10 items per category may have produced overlearning of the category names; thus the availability of category names as cues was the same for both whole-list recall conditions. The high rate of category recall for the Noncued and Cued Recall groups supports such a view.

## EXPERIMENT II

While the outcome of the first experiment strongly indicates that partial recall enhances within-category recall, a discrepancy in recall time between the Cued Partial Recall condition and the whole-list recall conditions may have produced the difference in recall. All three experimental conditions of Exp. I were given 15 min. for recall. However, the Partial Recall condition had only 10 words to recall in the 15 min., while the whole-list recall conditions had to recall from 100 words in the 15-min recall period. Experiment II attempted to equalize the time per word to-be-recalled for partial and whole-list recall conditions. In addition, conditions of input were manipulated; the words of a category were presented either in category blocks or randomly arranged in the word list, the list was presented only once, and the items per category were reduced to five.

### *Method*

*Design.*—Two presentation orders (blocked and random) and 3 cuing conditions (NonCued Recall, Cued Recall, and Partial Recall) were the factors of a 2 × 3 factorial design. The Noncued Recall and the Cued Recall groups were treated as in Exp. I. The Partial Recall group differed from that of Exp. I by requiring *S*s to recall all 10 categories of the list. All 3 recall conditions had 10 min. to recall 50 words; the Partial Recall condition was restricted to recalling one 5-word category per successive minute. The first category recalled was randomly chosen from the 10 categories of the list; the second category recalled was randomly chosen from the remaining 9 categories, and so forth, until all 10 cateogries had been cued for recall. It should be clear that the first category cued in the Partial Recall condition is the same as the Partial Recall condition of Exp. I, but with the important change that in Exp. II, the nominal recall time per to-be-recalled item was equated. The order of categories cued for the Partial Recall groups was matched across the blocked and random presentation conditions.

*Subjects.*—The *S*s came from the same *S* pool as in Exp. I. Fifty *S*s were tested, but 2 were rejected; 1 Partial Recall *S* failed to recall from a single category at a time, and 1 Noncued Recall *S* was a foreign student and was judged not to have an adequate facility with the English language. The 48 *S*s were equally and randomly distributed across the six cells of the 2 × 3 design; during the course of testing, the difference in number of *S*s between any two groups was never more than one.

*Word list.*—Five words were drawn from each 10-word category of Exp. I. The 50 words were either randomly assigned to input positions or they were blocked into categories which were then randomly ordered. The words were put on slides.

*Procedure.*—The *S*s were tested in groups of two or three persons. The *E* first read preliminary free recall instructions which were similar to those of Exp. I. The words were shown via slides at a 2 sec/word rate. After word presentation, *S*s read the instructions which differentiated the three experimental groups. The *S*s were allowed 30 sec. for instruction reading. The instructions were included in a booklet containing 10 pages of recall blanks. The Noncued Recall condition was simply given free recall instructions. The Cued Recall group was instructed to use the category names which appeared at the top of each recall page as aids in recall. The Partial Recall condition *S*s were required to recall from the category specified at the top of the recall page; each page had a different category name. All *S*s had 10 min. for recall. As *E* counted off the minutes, *S*s turned to a new page at the end of each minute. The *S*s in the Partial Recall condition recalled from a different category on each recall page; the *S*s of the whole-list recall conditions merely continued recall on all 10 pages.

### *Results and Discussion*

The data of the Partial Recall condition will be considered in two ways: (*a*) Performance on the first cued category will be compared with that under the whole-list recall conditions, Cued and Noncued Recall; the first cued category of the Partial Recall condition in this experiment is analogous to the Partial Recall condition of Exp. I. (*b*) The performance of the Partial Recall condition over all 10 categories will also be compared with the whole-list recall conditions. Table 1 presents a summary of the results.

The comparisons involving the first cued category of the Partial Recall condition will be presented first. When the mean words per category recalled values were compared, the partial recall effect of Exp. I was confirmed. Within-category recall to the Partial Recall condition. This depression of performance may be due to interference factors. Since the whole-list recall conditions were required to recall from more categories than the Partial Recall condition, the additional recall may have produced output interference. Whereas previous investigations of output interference (e.g., Murdock, 1963; Tulving & Arbuckle, 1963) have been limited to short-term memory situations, the present experiments show that output interference also occurs in secondary memory. The effects of the output interference was manifest in within-category recall; recall of a previous category reduced the amount of information recalled from the subsequent categories. The data of the successive recall of single categories in Exp. II suggest that output interference reached maximum after a single cued category. Smith et al. (1970), also using a successive category recall method, found the greatest within-category recall difference to occur between the first and second categories recalled. Further, the output interference studies of Murdock (1963) and Tulving and Arbuckle (1963) similarly showed that the output interference effect was limited to the output of one or two items.

Instead of output interference depressing the whole-list recall performance, a second type of mechanism is plausible. The partial recall effect may be viewed as due to facilitation of recall performance in the single category recall conditions. Atkinson and Shiffrin (1968) have suggested that control processes may enhance secondary memory search. The requirement of recalling from a single category in the partial recall condition may have resulted in a more efficient memory search by directing the search process or by providing *S* with an explicit category from which to recall. Partial recall may enhance performance in another way; it may allow *S* to exclude from memory the categories not cued, thus reducing interference from other items in the list (cf. Epstein, 1970). This would be a strategy which the whole-list recall conditions could not use. As with the output interference mechanism, the results of Exp. II would also place limits on the facilitation mechanisms. Facilitation by being the only category cued would be limited to the first category recalled. While the experimental results are consistent with interference and facilitation mechanisms, neither is dictated by the outcomes. It is perhaps prudent to consider both types of mechanisms to operate to produce the partial recall effect.

However, the partial recall effect does emphasize two important aspects of human verbal memory. First, it helps to stress the important role of retrieval processes in memory. Tulving (1969) has pointed out the value of the concept of retrieval cues in accounting for various memory phenomena. In addition to retrieval cues, the present study showed that conditions of recall also influence the amount of information recalled. Second, this study focuses attention on within-category recall. Since Miller (1956) suggested the importance of the chunk for understanding memory, much attention has been given to studying the relationship among chunks (e.g., Mandler, 1967). Interest in the content of chunks, on the other hand, has been limited to showing that within-chunk recall is secondary to chunk recall (Cohen, 1966; Tulving & Pearlstone, 1966). Because the partial recall method can specify *S*'s recall of any particular category or combination of categories, it may be a useful technique in investigating within-category recall.

## REFERENCES

Atkinson, R. D., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

Cofer, C. N., Bruce, D. R., & Reicher, G. M. Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 1966, **71**, 858–866.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 182–187.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms of verbal items in 43 categories. (Tech. Rep. No. 22, Nonr-631 (00)) Storrs, Conn.: University of Connecticut, 1957.

Epstein, W. Facilitation of retrieval resulting from postinput exclusion of part of the input. *Journal of Experimental Psychology*, 1970, **86**, 190–195.

Lewis, M. Q. Categorized lists and cued recall. *Journal of Experimental Psychology*, 1971, **87**, 129–131.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Murdock, B. B., Jr. Interpolated recall in short-term memory. *Journal of Experimental Psychology*, 1963, **66**, 525–532.

SHUELL, T. J. Clustering and organization in free recall. *Psychological Bulletin*, 1969, **72**, 353–374.

SMITH, A. D., D'AGOSTINO, R. R., & REID, L. S. Output interference in long-term memory. *Canadian Journal of Psychology*, 1970, **24**, 85–89.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

TULVING, E. Forgetting: Another look at an old problem. Paper presented at the Meeting of the International Congress of Psychology. London, July 1969.

TULVING, E. & ARBUCKLE, T. Y. Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 321–334.

TULVING, E. & OSLER, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593–601.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

(Received August 18, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 130–137

# SOME HYPOTHESES ABOUT NEGATIVE INSTANCES IN SINGLE-ATTRIBUTE CONCEPT ATTAINMENT[1]

STEPHEN V. HEIM AND ELLIN KOFSKY SCHOLNICK[2]

*University of Maryland*

Positive versus negative instances, number of binary dimensions (three vs. five), and the number of values changed per trial (one vs. two) were compared in single-attribute concept attainment. College students judged the relevance of each cue on each trial and received only sufficient information for solution. Initially, positive instance performance was superior, but it fell below negative instance performance by the final trial. Positive instance errors arose more often from memory deficits and from choice of irrelevant dimensions. Negative instance errors arose when interpreting information about present stimuli and the relevant dimension. Results were interpreted through a logical analysis of the task.

Most studies of adult concept attainment use conjunctive rules where values from at least two dimensions must be present simultaneously in the stimulus, e.g., red and square. The greater difficulty in identifying conjunctive concepts with negative as opposed to positive instances has prompted several informational explanations. First, if the dimensions are not binary, a negative instance eliminates fewer hypotheses than a positive instance (Hovland, 1952). If the choice is between red, blue, or green squares, a red negative instance leaves two possible alternatives, but a red positive instance leaves only one possibility. Even when the number of values eliminated with each instance is controlled, negative instance tasks still remain more difficult (Glanzer, Huttenlocher, & Clark, 1963; Huttenlocher, 1962, 1964). Hence differences in information rate alone do not account for the divergence in performance.

Second, it has been suggested that people are less accustomed to using negative instances. Although practice with negative exemplars quickly closes the performance gap (Freibergs & Tulving, 1961), the closure is not complete and the data do not explain the source of the original lack of practice.

There is evidence to support a third possibility. Processing negative information may require more inferential steps. Huttenlocher (1962, 1964) divides concept identification into phases of dimension location and value specification. In both positive and negative instance tasks, the learner locates the relevant dimension by finding the values which are common to all the stimuli he has seen. In the positive instance case, these values define both the relevant dimension and the values on it. With negative instances, the learner must convert those values of what the stimuli are *not* into positive instances by recalling the unseen alternatives on the relevant dimension. Young children often fail to perform the final step of value location (Scholnick, 1970).

Finally, there are two memory load explanations. First, in negative instance tasks, *S* eliminates hypotheses from a large set (Brayley, 1963; Hovland & Weiss, 1953). The learner defines all possibilities and remembers all disconfirmed hypotheses or those hypotheses which remain to be tested. In contrast, since the first positive instance defines all the hypotheses which could be correct, the burden on memory in positive instance tasks is less. An alternative memory load explanation advances the notion of category homogeneity (Bourne & Guy, 1968). Suppose *S* groups stimuli into either negative or positive instances. In conjunctive concept identification, there is greater variety among negative instances than among positive ones. If the concept is red squares, the negative instance class will contain blue squares and blue or red circles. In disjunction (either, or both), the situation is reversed. If the rule is red and/or square, red squares, red circles, and blue squares are positive instances and the sole negative instance is blue circle. Disjunctive rules produce fewer errors than conjunction under negative instance conditions while the reverse is the case for positive instances. But category homogeneity may also be equated by using single-attribute concepts and binary dimensions. If the concept is red, then all the negative instances will be blue.

---

[1] This article was based on a thesis submitted by the senior author in partial fulfillment of the requirements for the MA degree. The authors wish to thank Nancy Anderson, Thomas Turnage, and Janet Johnson for their suggestions. The research was supported in part by a grant from the Biomedical Science Support Committee of the University of Maryland to the Center for Language and Cognition and by a grant from the Computer Science Center, University of Maryland.

[2] Requests for reprints should be sent to Ellin K. Scholnick, Department of Psychology, University of Maryland, College Park, Maryland 20742.

The following experiment explored the roles of inferential complexity and memory load on single-attribute concept attainment where category homogeneity is controlled. The effects of memory load were examined by varying the number of irrelevant dimensions (Archer, Bourne, & Brown, 1955) and the rate at which hypotheses could be eliminated (Glanzer et al., 1963). It was predicted that increases in both these variables should affect negative instance concept attainment more than positive instance concept attainment. The experimental procedure required *S* to evaluate each of the possible stimulus cues on each trial to determine its relevance. These data permitted analysis of dimension and value location errors (Scholnick, 1970) and perceptual, inference, and memory errors by a scheme extended from the work of Cahill and Hovland (1960). As the memory load increased, the proportion of memory and inferential errors should also increase, particularly in the negative instance conditions.

## METHOD

*Stimuli.*—The concept problems were constructed from two stimulus sets differing in the number of component binary dimensions. One set was generated from three binary dimensions: form (circle or square), color (red or green), and size (271.0 sq. cm. or 899.4 sq. cm.). The five-dimension set contained additional variations in surface (striped or solid) and number of figures (one or two).

TABLE 1

STIMULUS SEQUENCES IN POSITIVE INSTANCE TASKS

| No. of stimulus dimensions | Information rate | |
|---|---|---|
| | 1 | 2 |
| 3 | small red circle<br>small red square<br>small green square | small red circle<br>small green square |
| 5 | 1 small solid red circle<br>1 small solid red square<br>1 small solid green square<br>1 small striped green square<br>2 small striped green squares | 1 small solid red circle<br>1 small solid green square<br>2 small striped green squares |

Note.—For the negative instances, large stimuli were substituted for small ones.

Four inferential problems varying in rate and type of stimulus information were constructed from each set. Each problem sequence provided sufficient information for identification of the relevant cue, small, after all the instances were shown. Each stimulus in one sequence differed from its predecessor in a single value. By deleting alternate stimuli, the stimuli in the second sequence differed from their immediate predecessors in two values. Within each of these conditions, one set of stimuli, all small, was labeled positive instances. The second set, all large, was labeled negative instances. Thus there were eight problems representing all combinations of two levels of complexity (5 vs. 3 bits), two types of instance (positive and negative), and two information rates (two vs. one). Hereafter each condition will be designated by three symbols listing the variables in the preceding order. Table 1 presents the four problems with all positive instances.

*Subjects.*—The *S*s were 298 students enrolled in an introductory psychology course (mean age = 20.1 yr.). Two were eliminated for failure on a color blindness test. Students were assigned to each condition according to their class schedule. Each task was administered to groups of 9 to 18 depending on their availability. There was an uneven number of *S*s per group ranging from 32 to 42 *S*s, with a minimum of 16 males and 16 females per group.

*Procedure.*—Each *S* performed two sample tasks prior to the experimental one. In all three, *S* had to determine the relevance of a list of possible cues by using information about a sequence of stimuli. A set of slides was projected one at a time for 30 sec. on a screen. Each slide contained a stimulus and either a plus or a minus sign designating whether the stimulus did or did not contain the relevant cue. While the slide was projected, *S* labeled each of the stimulus alternatives relevant or irrelevant on the

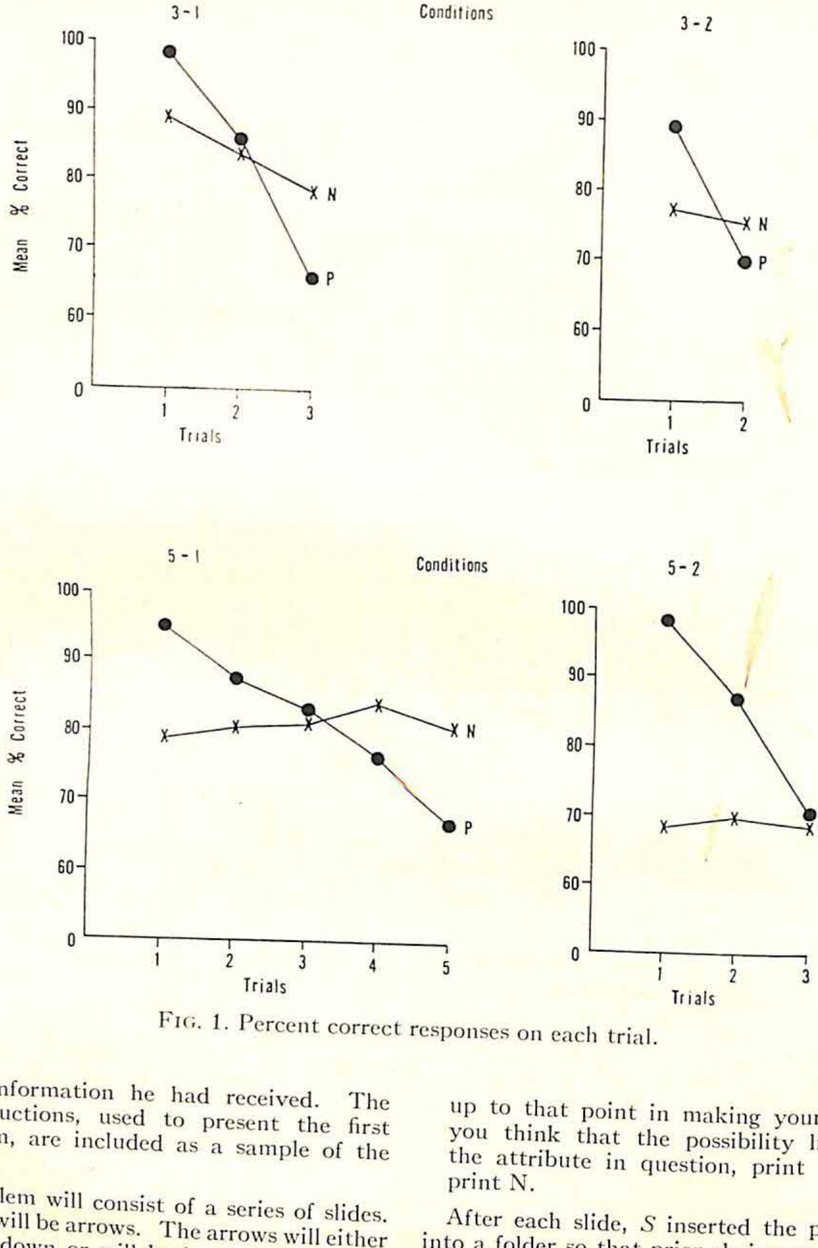


FIG. 1. Percent correct responses on each trial.

basis of the information he had received. The following instructions, used to present the first sample problem, are included as a sample of the procedure.

> Each problem will consist of a series of slides. The first set will be arrows. The arrows will either point up or down or will be located on the left or the right. I am thinking of one of those attributes, pointing up, pointing down, left side or right side and your job is to guess which one it is. I will give you some information that will help. If the slide I present contains the one cue I am thinking of, a plus sign will appear, and if it does not contain the cue, a minus sign will appear. In the booklet in front of you all the possibilities are listed on each page. After the presentation of a slide, indicate whether each of the possibilities listed could or could not be the one I am thinking of . . . Please note that the problem contains several slides so that you will have to use all the information you have received up to that point in making your decisions. If you think that the possibility listed could be the attribute in question, print Y and if not, print N.

After each slide, *S* inserted the page of choices into a folder so that prior choices were unavailable to *S*. On the first sample problem, *E* informed *S*s of the logical implications of positive and negative instances and of the correct answer. On the second task, *E* supplied only the correct answer upon completion. Both sample tasks contained a positive and a negative instance. Then the experimental task was presented. First *S* saw slides contrasting the stimulus values, then the problem slides were presented. Information about the solution was not supplied.

*Experimental design.*—There were eight groups of *S*s representing each combination of the three variables, instance type (positive vs. negative), number of stimulus dimensions (three vs. five), and stimulus values changing between trials (one vs. two). The cells contained unequal numbers of *S*s.

## RESULTS

Judgments of the relevance of each of the cues on each of the trials were converted into percentages of correct responses on the basis of information supplied. If *S* saw a large red circle as a positive instance on the first trial, large, red, and circle must be marked Y and small, green, and square, N. Each deviance was counted an error. If the second instance were a large green circle, now only red and circle must be marked Y and the remainder must be marked N. These scores were tallied for the sample task and each experimental trial. The data were converted to percentages since there were more judgments on the five- than the three-dimension tasks. Scores on the last trial of the *sample* task did not differentiate the experimental groups. Figure 1 shows the average percentages of correct responses on each trial of the eight experimental tasks. Because the tasks differed in the number of stimuli which had to be exposed before *S* had received sufficient information to solve the problem, initially only data from the last trial were analyzed. Just one variable, instance type, produced a significant difference in performance, $F$ (1, 278) = 6.11, $p < .02$, while rate and complexity of information and sex of *S* did not. Judgments were more accurate on the last trial with negative ($M = 80.3\%$) than with positive (73.6%) instances. There were also more *S*s who had identified the concept on the last trial in the negative instance conditions (49%) than in the positive instance tasks, 36%, $\chi^2$ (1) = 4.79, $p < .05$.

But the last trial analysis is somewhat misleading. In order to analyze a wider range of data, the scores for each of the four contrasting positive and negative instance conditions were compared over all trials, i.e., 3P1 versus 3N1, 3P2 versus 3N2, etc. In each of the four analyses, the effects of trials and type of instance were evaluated. In three of the analyses of variance, overall performance on the positive instance tasks was equal to performance with negative instances, but in the 5–2 comparison, performance with positive instances was more accurate, $F$ (1, 67) = 7.62, $p < .01$. In each condition, the course of performance with negative instances differed from performance with positive instances ($p < .01$). The trial-by-instance $F$ ratios for the 3–1, 3–2, 5–1, and 5–2 conditions were, respectively, $F$ (2, 148) = 7.18; $F$ (1, 74) = 9.25; $F$ (4, 284) = 15.34; and $F$ (2, 134) = 9.41. Performance on the initial trial with positive instances was almost perfect and significantly better than negative instance scores. But performance on positive instance tasks declined over trials, while negative instance performance remained constant except in the 3N1 condition, where it, too, declined.

*Dimensional relevance.*—Two different error analyses were performed to determine the factors contributing to performance. Errors were first categorized by dimensional relevance. If all the information *S* received enabled him to label the dimension irrelevant, but *S* did not, irrelevant dimension errors were scored. Conversely, if *S* labeled a dimension irrelevant when there was no information to justify its exclusion, relevant dimension or value location errors were scored. For example, after presentation of two positive exemplars, small red circle and small red square, misjudgments about the form dimension were scored irrelevant dimension errors and mistakes on the remaining dimensions were called relevant dimension errors. Errors were summed across all trials for all *S*s within a specific experimental task. Since there were differential opportunities for making each type of error, the total was divided by the number of opportunities. Table 2 presents the resulting percentages. Tests of the significance of the differences between percentages (McNemar, 1962) were carried out between groups who differed on instance type. In each comparison, there was a higher percentage of relevant dimension errors when negative instances were presented ($p < .001$) and a greater percentage of irrelevant dimension errors with positive instances ($p < .01$) except in the 5–2 task. Within groups, every positive instance condition produced

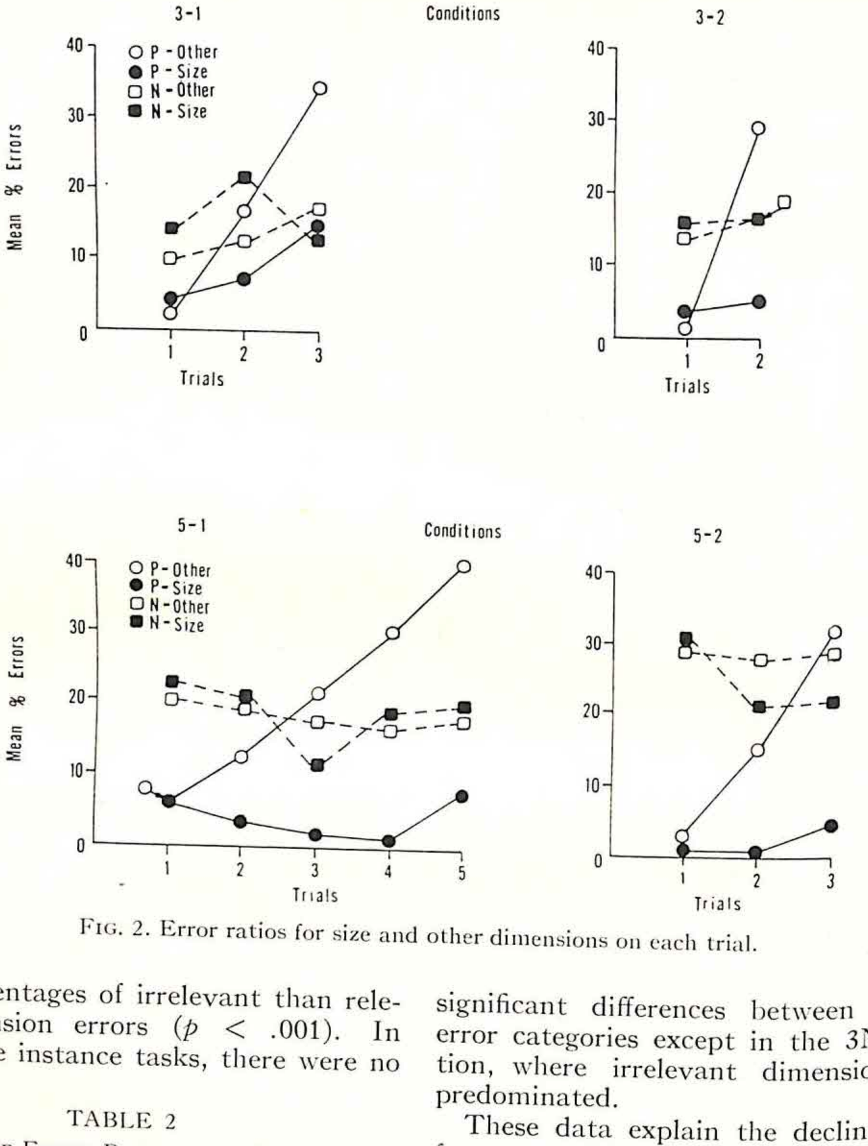


FIG. 2. Error ratios for size and other dimensions on each trial.

higher percentages of irrelevant than relevant dimension errors ($p < .001$). In the negative instance tasks, there were no significant differences between the two error categories except in the 3N1 condition, where irrelevant dimension errors predominated.

TABLE 2

COMPARISON OF ERROR RATIOS WITH POSITIVE AND NEGATIVE INSTANCES BY RELEVANCE

| Error | Cond. | | | |
|---|---|---|---|---|
| | 3-1 | 5-1 | 3-2 | 5-2 |
| Relevant | | | | |
| P | 4.9 | 3.0 | 3.3 | 1.7 |
| N | 11.9 | 18.9 | 14.4 | 27.2 |
| $z^a$ | −3.72* | −11.49* | −5.06* | −12.81* |
| Irrelevant | | | | |
| P | 32.8 | 21.6 | 30.4 | 30.2 |
| N | 18.2 | 16.0 | 16.1 | 30.6 |
| $z$ | +3.59* | +2.73* | +2.88* | −.12 |

[a] Each analysis contrasts the percentages of errors over error opportunities for positive versus negative instance conditions.

* $p < .01$.

These data explain the decline in performance with positive instances and the stability of negative instance performance. The later the trial, the more hypotheses must be eliminated and the more irrelevant dimension errors contribute to performance. On the first trial, *S* can eliminate one value on each dimension but no dimension itself. On the last trial of the 5–1 task, four dimensions or four-fifths of the errors will be contributed by the irrelevant dimensions. In the positive instance conditions, where irrelevant dimensions exceeded relevant errors, irrelevant errors contributed more and more to performance. In the negative instance cases, where errors from the two sources were equal, performance was stable.

Because irrelevant dimension errors cannot be produced on the first trial and opportunities for irrelevant dimension errors varied widely among the trials, learning curves would be difficult to assess. Instead, an alternative category system was developed. Figure 2 presents the errors categorized by size, the predesignated relevant dimension, and by those dimensions which will eventually become irrelevant as more information is received. As before, the number of errors committed was weighted to equalize the number of opportunities for committing errors. Eight analyses of variance were performed contrasting trial and error categories within a given task. The trial effects were reported earlier. In three of the four positive instance conditions, 3–2, 5–1, and 5–2, size errors were less frequently produced than errors in the other category. The $F$ ratios for the three conditions were, respectively, $F(1, 41) = 8.92$, $F(1, 124) = 35.95$, and $F(1, 68) = 30.40$, $p < .01$. While size errors remained constant across trials in each positive instance condition, the proportion of erroneous responses to the other dimensions increased on each trial; hence the two error categories diverged increasingly on later trials. The four significant Category × Trial interaction effects in the 3–1, 3–2, 5–1, and 5–2 conditions were, $F(2, 66) = 12.88$; $F(1, 41) = 25.94$; $F(4, 124) = 15.81$; $F(2, 68) = 10.25$, $p < .01$. In the negative instance tasks, each error category produced the same proportion of errors and the error rate remained constant across trials.

*Memory errors.*—Errors were also divided into perceptual, inference, and memory categories. A perceptual error occurred when the cue was present on a given trial but *S* misinterpreted information about it. If the positive instance was a small circle, but *S* judged small or circle to be irrelevant, a perceptual error was scored. When a value was absent from the stimulus figure on that trial, but its relevance could be assessed from a value which was present, incorrect judgments were scored inference errors. Designating large or square as relevant constituted an inference error. Finally, a memory error involved a decision which had to consider previously shown stimuli in order to be correct because the stimulus on the present trial differed on a dimension from preceding stimuli. When a small square followed a small circle, *S* must eliminate both forms on the second trial since both cannot be relevant.

TABLE 3

Comparison of Error Ratios with Positive and Negative Instances—Perceptual, Memory, and Inference Data

| Error | Cond. | | | |
|---|---|---|---|---|
| | 3-1 | 5-1 | 3-2 | 5-2 |
| Perceptual | | | | |
| P | 5.4 | 4.8 | 4.8 | 1.6 |
| N | 8.7 | 17.9 | 10.8 | 23.2 |
| $z^a$ | −1.45 | −6.91* | −2.12* | −11.53* |
| Inference | | | | |
| P | 6.5 | 8.1 | 4.8 | 5.5 |
| N | 16.3 | 17.9 | 14.7 | 24.8 |
| $z$ | −3.67* | −5.58* | −3.40* | −8.08* |
| Memory | | | | |
| P | 52.9 | 41.6 | 47.6 | 52.4 |
| N | 29.4 | 18.3 | 26.5 | 41.2 |
| $z$ | +3.62* | +6.91* | +2.67* | +2.18* |

[a] Each analysis contrasts the percentages of errors over error opportunities for positive versus negative instance conditions.

* $p < .05$.

Table 3 presents the percentages of each error type which were computed and compared in the same manner as the relevant-irrelevant analysis. Only contrasts in instance type produced consistently significant differences. There were more perceptual and inference errors with negative than with positive instances in three of the four comparisons. Conversely, in all tasks, memory errors were higher with positive than with negative examples. Memory errors contributed most to the total error scores since their frequency was greater than the other two categories in all but the 5N1 condition. Trial analyses were not possible since memory errors cannot occur on the first trial.

## Discussion

If processing negative instances places a greater burden on memory than processing positive instances, then the performance gap between negative and positive instance tasks should widen as the number of irrelevant dimensions increases or as the number of stimulus values which change between trials increases. But these predictions were not confirmed. Neither rate of change nor task complexity affected performance, but the third variable, information type, produced differences in accuracy. Each instance type elicited a different distribution of error types and different learning curves. Initially, inference from positive instances was more accurate than inference from negative instances but that initial superiority rapidly vanished. Inferential errors increased over trials on positive instance tasks while negative instance performance remained constant. When *S*s could be correct by simply attending to the instance in front of them, *S*s assigned to positive instance tasks made fewer errors, as Tables 2 and 3 show. Additionally, *S*s made fewer value specification errors on positive instance tasks. These data confirm Huttenlocher's (1964) claims concerning the greater difficulty of identifying the correct value on the relevant dimension when exposed to negative instances.

Huttenlocher (1964) also predicted that dimension location errors would be equally frequent in both conditions. Both tasks produced many irrelevant dimension choices but fewer occurred in the negative instance tasks. In addition, past experimental data suggested that memory errors would be more prominent in the processing of negative information but again fewer were found with negative instances. Since the percentage of decisions involving irrelevant dimensions and use of memory increased across trials, performance on the last trial was more accurate with negative information than with positive.

Processing of past data from positive instances may be logically more complex because positive instances must be transformed into negative ones. If the first positive instance is red and the second green, two positive instances make a negative or irrelevant dimension. Conversion of information is unnecessary with negative instances. If red is eliminated with the first negative instance, a negative green instance simply erases the second value on the dimension. The decline in performance for positive instance tasks can be explained using the analysis of memory errors. Both irrelevant dimension and memory judgments require conversions of two positives into a negative. These decisions increase on later trials. Neither memory nor irrelevant dimension errors are possible on the first trial; on the last trial, all but one dimension must be eliminated, and memory decisions account for over a third of the choices. Positive instance *S*s rarely make any other kind of error, and their performance declines as the problem progresses.

A more global explanation of the data posits an ideal strategy for each task. In positive instance problems, *S* must select, as relevant, values present in the stimulus and must reject absent values. The strategy appears to be adopted since few perceptual and inference errors occur. When only squares appear, square is rarely judged irrelevant and circle relevant. The sole exception to the strategy arises when *S* has past information which compels rejection of some of the values present. Then, *S*s often fail to convert two positive instances on the same dimension into negatives. The ideal strategy for negative instances is the opposite of the positive instance strategy and requires no exceptions. Whatever values are seen are irrelevant and absent values are relevant. Because past instances fail to violate this general exclusion strategy and no rule changes are necessary, memory decisions involving negative instances are logically less complex. Memory errors should occur because *S*s forget what they have seen or because people fail to adopt the exclusion strategy. The failure to adopt the strategy is documented by the high proportion of perception and inference errors. The greater frequency of these errors in the negative instance tasks supports previously reported data (e.g., Dominowski, 1968) on the difficulty of transforming negative into positive instances. The high frequency of perception errors suggests that rather than avoiding conversion of information, often *S*s inappropriately converted negative into positive instances. Thus negative green instances were treated as if they were positive. In summary, the exclusion strategy is simpler using negative instances, but fewer people adopt the strategy in negative instance than positive instance tasks.

The preceding explanations stem from analysis of single-attribute concept tasks. The data generally used to compare positive and negative instances differ from the current data in several respects. The present experiment suggests that the relative superiority of positive or negative information varies according to the number of previous trials and the error category. Most conventional analyses ignore irrelevant dimension errors because the experimental procedure does not permit determination of the source of errors. If just errors on the size dimension were computed the present data would resemble the traditional analyses of the effects of negative instances in concept learning. Additionally, if the learner is given just sufficient data for solution and must respond to all aspects of the stimuli, he must make hypotheses simultaneously rather than successively. He must resort to an exclusion strategy. Thus the demands of the task are quite different from conventional concept attainment tasks. The task differs in a third way. When single-attribute concepts are used, the homogeneity of positive and negative instance classes is equated. Conjunctive concepts, usually employed with adults, lack this feature. Further experimental work should clarify which task variable, homogeneity of instances or the imposition of an exclusion strategy, is responsible for the rapid decline in processing positive instances and the stability in negative instance performance. Additionally, more complex problems, requiring more trials for solution, would provide further data on the relative incidence of memory errors in positive and negative instance tasks. The current test of the memory load hypothesis was not optimal. Although increased stimulus complexity produced poorer performance the difference failed to reach a statistically significant level. Most of the complexity effects have occurred with mixed instances and widely separated numbers of dimensions (Archer, Bourne, & Brown, 1955; Bourne, 1957; Brown & Archer, 1956; Pishkin, 1960, 1965).

## REFERENCES

ARCHER, E., BOURNE, L., & BROWN, F. Concept identification and irrelevant information. *Journal of Experimental Psychology*, 1955, **49**, 153–164.

BOURNE, L. Effects of delay of information feedback and task complexity on identification of concepts. *Journal of Experimental Psychology*, 1957, **54**, 201–207.

BOURNE, L., & GUY, D. Learning conceptual rules. II: The role of positive and negative instances. *Journal of Experimental Psychology*, 1968, **77**, 488–494.

BRAYLEY, L. Strategy selection and negative instances in concept learning. *Journal of Educational Psychology*, 1963, **54**, 154–159.

BROWN, F., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *Journal of Experimental Psychology*, 1956, **52**, 316–321.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, **59**, 137–144.

DOMINOWSKI, R. L. Stimulus memory in concept attainment. *Psychonomic Science*, 1968, **10**, 359–360.

FREIBERGS, V., & TULVING, E. The effect of practice on utilization of information from positive and negative instances in concept formation. *Canadian Journal of Psychology*, 1961, **15**, 101–106.

GLANZER, M., HUTTENLOCHER, J., & CLARK, W. Systematic operation in solving concept problems: A parametric study of a class of problems. *Psychological Monographs*, 1963, **77** (2, Whole No. 564).

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychological Review*, 1952, **59**, 461–472.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology*, 1953, **3**, 175–182.

HUTTENLOCHER, J. Some effects of negative instances on the formation of simple concepts. *Psychological Reports*, 1962, **11**, 35–42.

HUTTENLOCHER, J. Development of formal reasoning on concept formation problems. *Child Development*, 1964, **35**, 1233–1242.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1962.

OSLER, S. F., & KOFSKY, E. Stimulus uncertainty as a variable in the development of conceptual ability. *Journal of Experimental Child Psychology*, 1965, **2**, 264–279.

PISHKIN, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *Journal of Experimental Psychology*, 1960, **59**, 371–378.

PISHKIN, V. Dimension availability with antecedent success or failure in concept identification. *Psychonomic Science*, 1965, **2**, 69–70.

SCHOLNICK, E. K. Inference and preference in children's conceptual performance. *Child Development*, 1970, **41**, 447–460.

(Received August 18, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 138–142

# REACTION TIME UNDER STIMULUS UNCERTAINTY WITH RESPONSE CERTAINTY[1]

PAUL J. BARBER[2]
*Birkbeck College, University of London*

SIMON FOLKARD
*Medical Research Council*
*Developmental Psychology Unit, London*

An effect of stimulus set size was found with a paradigm in which all stimuli were mapped onto a single response. The effect was essentially unchanged when the stimuli (letters) were inverted or when each stimulus was also associated with a unique foveal location. The effect was found to be transient, and the results may be interpreted as indicating a switch in processing strategy from a focal attentive to a preattentive mode.

Early studies of the relationship between choice reaction time (choice RT) and stimulus uncertainty (Hick, 1952; Hyman, 1953) used 1:1 stimulus-response mappings and thus confounded stimulus and response uncertainty (Smith, 1968). Judicious manipulation of many:1 mappings (Crossman, 1953; Fitts & Biederman, 1965; Rabbitt, 1959) provides one way of effecting the unconfounding of these two variables. A related way of examining the role of stimulus uncertainty is to manipulate stimulus set size ($s$) with a total response ensemble size of $r = 1$. Thus in this paradigm all stimuli are mapped onto a single response.

Bernstein, Schurman, and Forester (1967) reported a pertinent study in which *S*s responded with set sizes $s = 1$, 2, 4, and 8, with a single response $r = 1$. The resulting monotonic increase in choice RT with $s$ is equivocal with regard to the role of stimulus uncertainty since *S* was additionally asked to guess which stimulus was to be presented before each trial. If *S* tested the accuracy of his guess before initiating his motor response and if hypothesis checking of this sort took longer with larger $s$, then choice RT would increase as it did. At the very least the response demands of the task were not straightforward. Furthermore since *S*s were also tested under conditions where $r > 1$, treatment carryover phenomena could in part have mediated the observed effect of $s$.

Hannes, Sutton, and Zubin (1968) also reported a study in which stimulus uncertainty was associated with response certainty. Using a novel paradigm involving two signals (1 sec. apart), both requiring the same response, the response to the second signal was facilitated when the signal sequence involved no uncertainty. The effect here could reflect intermodal attentional adjustments brought about by the particular double-signal arrangement used.

In contrast, the most direct investigation using a response certainty paradigm seems to be that of Griew (1958) in which *S*s made a single response to spatially distributed signals ($s = 1$, 2, 4, and 8). Stimulus uncertainty was again effective (but differently in young and old *S*s). In this case explicit orienting responses may have been necessary for *S* to accommodate the full display. The emergence of an effect of uncertainty at a smaller value of $s$ for the older *S*s is congruent with this interpretation in view of the reduced visual field in old *S*s (Ferree, Rand, & Monroe, 1930).

The present study is an equally direct application of the all:1 paradigm but avoids the problem of gross attentional scatter. At the same time, it keeps the task load minimal and uses intramodal stimulus uncertainty. While this is an "obvious" way of manipulating stimulus uncertainty unconfounded by response uncertainty, and it is a formally valid method, it may be objected that it does not bear on the choice process *because* no degree of response choice is involved. The method is, of course, neutral with respect to this issue.

---

[1] This research was conducted while the second author was in receipt of Medical Research Council Grant G.77/1931. The authors are grateful to P. H. Venables for his comments on an earlier draft of this manuscript.

[2] Requests for reprints should be sent to Paul Barber, Department of Psychology, Birkbeck College, Malet Street, London, W. C. 1, England.

If only a single response is required, then it is logically unnecessary for *S* to make any stimulus categorization, beyond the level of classifying the input as "stimulus present" or "stimulus absent." Clearly if *S* operated thus, the all:1 paradigm would reduce to a simple RT paradigm. Certainly in the present case *S* can most efficiently perform by responding to stimulus onset (as evidenced by a brightness change). On the other hand, more complex categorization may in some sense be obligatory. Similarly, *S* may habitually extract or attempt to transmit more information than the situation demands. Nevertheless an effect of *s* would be masked if *S* adopted a simple RT strategy for his motor response and read out content information from iconic storage (Neisser, 1967) after the physical display had terminated.

In the present experiment, the display information could be considered as irrelevant. In the choice RT context, there is precedent for an expectation that irrelevant information can be ignored, at least in the sense of not modifying the effect of *s* (Morin, Forrin, & Archer, 1961). Obligatory categorization of irrelevant information has in contrast been observed by Biederman and Checkosky (1970), although it should be emphasized that the information was irrelevant by virtue of being redundant. Given that the display content is irrelevant in the present task, it would seem that the absence of an effect is the more likely here.

"Efficient" performance in this all:1 paradigm is achieved by ignoring content information. Letters were in fact used as stimuli, and since in some studies digit/letter-naming latencies have been found to be independent of *s* (Mowbray, 1960), it was thought appropriate to include an operation designed to bear on the naming process. Since letter recognition has been found to be sensitive to stimulus orientation (Henle, 1942; Robinson, Brown, & Hayes, 1964), the effect of stimulus inversion was examined here.

A further experimental manipulation was concerned with the effect of spatial uncertainty within foveal vision. Furthermore, within-sessions effects were examined using single RT data. Both Smith (1967) and Bartz (1971) reported an effect of *s* when first RTs were considered, but in both cases potential response uncertainty existed.

## METHOD

Stimulus set sizes of $s = 1$, 2, and 4 were presented under three conditions, with a response set size of $r = 1$ in all cases. Cond. A: The stimuli, alphabetic characters, were displayed with their normal orientation. All stimuli were presented at the same place on the display. Cond. B: As Cond. A, but the stimuli were rotated through 180°. Cond. C: As Cond. A, but the stimuli were in addition spatially separated, each being uniquely assigned to its intrafoveal location so that content and location were perfectly correlated.

*Subjects.*—The *S*s were 33 male and 39 female members (age range 18 to 39 yr.) of the departmental *S* panel. They were not generally naïve to psychological experiments but were naïve specifically with respect to reaction time studies. The *S*s were paid for their services at a rate of 30 pence (approximately 72¢) per hour.

*Apparatus and stimuli.*—The stimuli were alphabetic characters presented via the character-display facility of the oscilloscope (equipped with a P19 phosphor) of a LINC-8 computer. The character set consisted of the letters F, G, J, and K; each was formed within a 6 × 4 matrix of points measuring approximately .003 × .002 m. Each character required the intensification of 12 of the 24 possible points so that they were roughly equated for intensity. In Cond. A and B, the stimuli appeared at the center of an outline square frame of .02 × .02 m., drawn on the scope face as an aid to fixation. In Cond. C, the stimuli were centered at the four corners of a .005 × .005 m. square concentric with the fixation frame.

In all cases, *S* viewed the display from about .6 m. and responded via a .013-m. diameter button operating a microswitch. Ambient illumination was such that the phosphorescent afterglow of the display was essentially imperceptible.

The computer program monitored responses and recorded latencies in milliseconds. Response-stimulus intervals were pseudo-randomly generated rectangularly on the interval 1.512 to 2.535 sec. The program also rotated the order of presentation of stimulus set sizes between *S*s, generated pseudo-random permutations of the stimuli within blocks, controlled the display, and performed data tabulation and reduction operations.

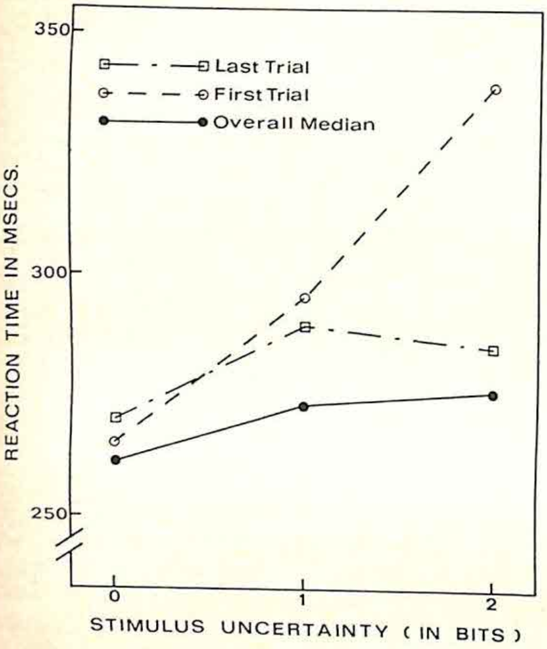


FIG. 1. The effect of stimulus uncertainty on reaction time in the all:1 paradigm. (Results are shown averaged across conditions (72 *S*s) for median reaction times and first and last reaction times.)

*Procedure and design.*—Written instructions were given to *S*. Besides informing him about general details of the procedure, the task and its requirements were explained as follows:

> This experiment is designed to find out how quickly you can react to a symbol on the screen. When a symbol appears you are to press the key as quickly as possible, but please do *not* press the key when there is no symbol displayed.

The procedure was further described as involving "several sessions" and lasting "about 20 minutes." The range of stimuli was displayed prior to each block until *S* pressed his key to initiate the stimulus presentations for that block. A buzzer sounded for .5 sec. at the end of each block.

The *S*s were allocated to three groups of 24, corresponding to Cond. A, B, and C. Within each group, 4 *S*s were tested under each of the six possible presentation orders of the three set sizes. All *S*s were tested at each level of $s$ in one block of 100 trials per level. For the $s = 2$ and $s = 4$ blocks, the corresponding stimuli were equally frequent and the order of presentation within blocks was generated for each *S* independently. The entire session lasted about 15 min. at most.

A retention test was given to the first 12 *S*s tested under Cond. A and B. They were provided with three rows of five $6 \times 4$ matrices with instructions to draw the experimental symbols for each block as accurately as possible. Only 3 *S*s out of 24 made as much as a single error of reconstruction, and none made an error of order. This part of the procedure was thereafter eliminated since a substantial change in this effectively perfect recall seemed most unlikely.

## RESULTS

All analyses were based on RTs to the symbol J, which appeared in all three blocks encountered by every *S*. Median RTs, used in the preliminary analysis of variance, were based on every fourth RT to J in the $s = 1$ block, every second RT to J in $s = 2$ block, and all RTs to J in $s = 4$ block. Thus all medians were based on 25 observations distributed regularly within each block.

In this first analysis, only the variance due to the within-*S*s factor, set size, was significant, $F$ (2, 108) $= 7.44$, $p < .001$. Both other main effects, the between-*S*s factors conditions and presentation order, were not significant, nor were any of the interaction terms ($p$s $> .25$ in all cases). Partitioning the variance attributable to set size yielded a highly significant linear component, $F$ (1, 108) $= 12.71$, $p < .001$, and nonsignificant deviation from linearity, $F$ (1, 108) $= 2.20$, $p > .10$. The option of using the equal interval property on the stimulus uncertainty variable was exercised in preference to the unequal intervals on the set size factor. The statistical conclusions are independent of this option in the present analyses.

This reliable overall effect of set size (Fig. 1) was explored further by an analysis of the first and last critical RTs. It should be emphasized that this second analysis cannot be said to be independent of the first one, but stands in qualification thereof. Set size remained significant, $F$ (2, 108) $= 7.998$, $p < .001$, but now the extra within-*S*s variable, trials (first vs. last presentation of J), interacted significantly with set size $F$ (2, 108) $= 4.052$, $p < .025$. No other main effect nor interaction was significant.

The Trials × Set Size interaction (Fig. 1) was more specifically found to be attributable to a within-block erosion of the set size effect. Thus on the first trial, the linear component of set size was highly significant, $F$ (1, 108) $= 22.46$, $p < .001$ but on the last trial was not significant,

$F$ (1, 108) = 1.44, $p > .25$. This was further substantiated by separate analyses of first and last critical RTs, conducted as a precautionary measure due to the sizable decrease in the variability within blocks. In any event, the pattern of statistical significance is maintained even if ultra-conservative $F$ tests are used following Greenhouse and Geisser (1959).

It can be seen that in Fig. 1, means of median RTs are on the whole rather less than the means of the single RTs. This tendency seems unimportant and could merely be a statistical artifact attributable to the effects of positive skew in RT distributions, although it may reflect a within-blocks, U-shaped, sequence effect.

The absence of conditions effects could be due to criterion differences between groups, independent of the experimental treatments. For this reason, the number of *S*s making anticipatory responses, arbitrarily defined as latencies below 100 msec., was obtained. In Cond. A, B, and C there were, respectively, 14, 12, and 11 out of 24 *S*s making such responses. Anticipations were clearly not associated with conditions and trade off seems an unlikely explanation for the failure of the conditions factor to contribute significant variance alone or jointly with some other factor.

## Discussion

The experimental procedure used here guarantees that response uncertainty is zero in the formal sense of information theory. In this same sense we may infer from the results that stimulus uncertainty here has a reliable but short-lived effect when there is no response uncertainty. It seems, then, that the effect of *s* is not altogether a response effect. This is consistent with recent findings that stimulus probability has effects which are independent of response probability (Biederman & Zachary, 1970; Hawkins & Hosking, 1969).

It is nevertheless arguable that the issue in the present study is confounded by the use of letters as the stimulus ensemble and that the observed effect of *s* only reflects a tendency on the part of *S*s to name the letter before making the manual response. If this is the case, then *S*s evidently prefer to make use of distal information rather than some iconic residue, and at least in the short run implicit naming of a well-defined set of letters is itself a function of *s*. This contrasts with the findings of Mowbray (1960), but it is consistent with the small effect on letter naming observed by Morin, Konick, Troxell, and McPherson (1965).

In Neisser's (1967) terms, it appears that *S*s brought serial focal attentive processes to bear on a task for which parallel preattentive treatment would seem more suitable. That *S*s could so easily recall the stimuli further reinforces the notion that focal attention was operative. The disappearance of the effect of *s* within the session suggests a switch between processing strategies such that by the end of each block *S* had resorted to preattentive processing. It is as if *S* initially attended to the display content within a block, and hence categorized each stimulus. As soon as he was certain that the briefing display had correctly specified the contents of that block, he could adopt a simple RT strategy.

The failure of the letter inversion treatment (Cond. B) to have a differential effect mitigates against the results, reflecting a hidden response effect. This point rests on the absence of a statistical effect and the possibly naïve presumption that letter inversion would affect naming processes. Foveal spatial uncertainty (Cond. C) also seems to have added nothing to the effect of *s*. The point of this manipulation also was to allow for response effects; in this case, the notion was entertained that intrafoveal orienting processes might be capable of mediating an effect of *s*. The lack of a distinctive contribution from these manipulations is surprising and is at odds with a response-based interpretation. It would be inappropriate to overextend this point, but the two experimental treatments B and C are salient operations for testing of template-matching versus feature-testing issue in pattern recognition (Neisser, 1967).

The finding that *S*s do go beyond the information required, that they operate as what might be called information generators, is consistent with other evidence that irrelevant information is not always easy to ignore (Triesman, 1969). It then remains to be explained why irrelevant information should be processed. The *S* may, of course, have difficulty in filtering out irrelevant information, especially when attention has to be focused on a single analyzer or group of analyzers (Triesman, 1969). This seems unlikely in the present case where efficiency could be achieved by switching out the entire bank of analyzing devices responsible for display content.

Information generation could be the involuntary and obligatory consequence of a mechanism, presumably classificatory, extracting information beyond the organism's needs. But if the effect arises on a nonobligatory basis, and it is tempting partly because of the transience of the effect to credit *S*s with some semblance of autonomy, then it follows that *S*s may actively and spontaneously manipulate structural subprocesses in reaction time tasks.

## REFERENCES

Bartz, A. E. Reaction time as a function of stimulus uncertainty on a single trial. *Perception & Psychophysics*, 1971, **9**, 95–96.

Bernstein, I. H., Schurman, D. L., & Forester, G. Choice reaction time as a function of stimulus uncertainty, response uncertainty, and behavioral hypotheses. *Journal of Experimental Psychology*, 1967, **74**, 517–524.

Biederman, I., & Checkosky, S. F. Processing redundant information. *Journal of Experimental Psychology*, 1970, **83**, 486–490.

Biederman, I., & Zachary, R. A. Stimulus versus response probability effects in choice reaction time. *Perception & Psychophysics*, 1970, **7**, 189–192.

Crossman, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice responses. *Quarterly Journal of Experimental Psychology*, 1953, **5**, 41–51.

Ferree, C. E., Rand, G., & Monroe, M. M. A study of the factors which cause individual differences in the size of the form-field. *American Journal of Psychology*, 1930, **42**, 63–71.

Fitts, P. M., & Biederman, I. S-R compatibility and information reduction. *Journal of Experimental Psychology*, 1965, **69**, 408–412.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. *Psychometrika*, 1959, **24**, 95–112.

Griew, S. Uncertainty as a determinant of performance in relation to age. *Gerontologia*, 1958, **2**, 284–289.

Hannes, M., Sutton, S., & Zubin, J. Reaction time: Stimulus uncertainty with response certainty. *The Journal of General Psychology*, 1968, **78**, 165–81.

Hawkins, H. L., & Hosking, K. Stimulus probability as a determinant of discrete choice reaction time. *Journal of Experimental Psychology*, 1969, **82**, 435–440.

Henle, M. An experimental investigation of past experience as a determinant of visual form perception. *Journal of Experimental Psychology*, 1942, **30**, 1–22.

Hick, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, **4**, 11–26.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, **45**, 188–196.

Morin, R. E., Forrin, B., & Archer, W. Information processing: The role of irrelevant stimulus information. *Journal of Experimental Psychology*, 1961, **61**, 89–96.

Morin, R. E., Konick, A., Troxell, N., & McPherson, S. Information and reaction time for "naming" responses. *Journal of Experimental Psychology*, 1965, **70**, 309–314.

Mowbray, G. H. Choice reaction times for skilled responses. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 193–202.

Neisser, U. *Cognitive psychology*, New York: Appleton-Century-Crofts, 1967.

Rabbitt, P. M. A. Effects of independent variations in stimulus and response probability. *Nature*, 1959, **183**, 1212.

Robinson, J. S., Brown, L. T., & Hayes, W. H. Tests of effects of past experience on perception. *Perceptual & Motor Skills*, 1964, **18**, 953–956.

Smith, E. E. Effects of familiarity on stimulus recognition and categorization. *Journal of Experimental Psychology*, 1967, **74**, 324–332.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

Triesman, A. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–289.

(Received August 19, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 143–150

# MANIPULATED RETRIEVABILITY IN FREE RECALL

ROBERT K. YOUNG[1] AND A. KEITH BARTON

*University of Texas at Austin*

Ease of free recall of responses following paired-associate learning was found to be directly related to degree of stimulus imagery in Exp. I. Two explanations, called the mediation hypothesis and the acquisition hypothesis, were used to account for these data. The former holds that recall of a response is mediated by its associated stimulus, and the latter holds that some attribute of the stimulus is acquired by the response associated with it. Experiment II investigated response recall as a function of the relatedness of the stimuli, and Exp. III investigated response recall as a function of stimulus imagery and transfer paradigm. The data were generally supportive of the acquisition hypothesis.

In a previous study, Young and Parker (1970) attempted to investigate the acquisition of meaningfulness (M) in the laboratory. In this study, *S* first learned a common set of low-M responses to one of three sets of paired-associate (PA) stimuli which varied in M and then learned the same set of responses in a serial list. The difficulty of the serial list was inversely related to the M of the stimuli to which the responses had been associated. These results supported the hypothesis that the M value of the responses was directly related to the M value of their associated stimuli because the stimuli served as mediators between the low-M responses and either a large (in a case of high M) or a small (in the case of low M) number of associates.

The first experiment in the present article (Exp. I) was designed to investigate the possibility that imagery (I) (Paivio, 1969) could be induced in the laboratory in a manner similar to that used by Young and Parker (1970). As in the previous study, the manipulated variable was significant: free recall of responses associated with high-I stimuli was greater than free recall of responses associated with low-I stimuli.

Based on the results of Exp. I, two explanations of these data were suggested. It may be that during PA learning the responses acquired some portion of the imagery possessed by the associated stimuli, or it may be that the critical variable in both of the previous studies was the availability of the stimuli. If the latter alternative is the case, variable stimulus recall is being manipulated rather than M or I.

Experiment II was designed to investigate the possibility that stimulus mediation plays an important role in this paradigm. This was done by investigating response learning as a function of stimulus retrievability with attributes of the stimulus materials such as meaningfulness and imagery held constant.

The results of Exp. II suggested that manipulation of the retrievability of stimuli influences free recall of the responses of a PA list. Because of this, acquired response I, variable stimulus recall, or both could account for the results obtained in Exp. I. Experiment III represents an attempt to differentiate between these two explanations of the results of Exp. I. This was done by having *S*s learn PA lists in an A-B, A-C paradigm. Following A-B learning, Response B is more easily recalled when A has high I than when A has low I. Under the mediation hypothesis, the ability of the stimuli to serve as mediators for recall of the B responses is reduced following A-C learning because the A-B associations have been unlearned. Once A-B unlearning occurs, the B responses previously associated with easily recalled (high-I) stimuli should show a greater retention loss relative to an appropriate control than should those responses previously associated with difficult-to-recall stimuli. The reason for this, of course, is that loss of a mediator which increased recall will result in a considerable retention loss, while loss of a mediator which did little to facilitate recall will result in only a relatively small loss.

---

[1] Requests for reprints should be sent to Robert K. Young, Department of Psychology, University of Texas at Austin, Texas 78712.

On the other hand, if Response B acquires I (or some component of I) during A-B learning, then unlearning of the A-B association following A-C learning should not reduce the differential recall of responses associated with high- or low-I stimuli. That is, even though the B responses are no longer associated with A, responses formerly associated with high-I stimuli should still be recalled better than responses formerly associated with low-I stimuli.

## EXPERIMENT I

### *Method*

*Design.*—Each *S* learned an eight-pair PA list for three or four trials and then learned the PA responses for five trials by free recall. For half of the *S*s, the stimuli of the PA list had high I. For the other half of the *S*s, the PA stimuli were low I. The responses for both lists consisted of a common set of low-M trigrams. Thus the design is a 2 × 2 with number of PA trials and degree of I being factorially combined ($N = 11$ per group).

*Procedure.*—The PA lists were learned by alternating observation trials in which both members of the pair were presented and test trials in which only the stimuli were presented. Both the pairs and the stimuli were presented at a 3-sec. rate with a 9-sec. intertrial interval. Subsequent to PA learning, *S*s were given free recall instructions and were told to recall the responses ("those three-letter words you have been writing down") which they had learned during PA learning. During free recall, the items were presented at a 3-sec. rate, a 24-sec. recall interval, and a 9-sec. test trial–observation trial interval were used. Although observation and test trials were alternated, the first test trial came before the first observation trial in free recall. The *S*s were run in a vacant classroom and were seated in every third seat. The verbal materials were presented by a Kodak Carousel Ektagraphic projector. Recall was written, and after every trial *S* placed his data sheet face down on an adjacent desk and used another sheet for the next trial. The *S*s were run in groups of five or six. Four orders of presentation were used during both PA learning and free recall to reduce the possibility of serial learning.

Response M was low with the Archer (1960) ratings ranging from 19% to 23% association value with the median equal to 20. The stimuli were taken from norms developed by Paivio, Yuille, and Madigan (1968). The stimuli with high I had a mean rating of 6.33 and a range from 6.10 to 6.53; the stimuli with low I had a mean of 2.70 and a range from 2.50 to 2.93. Stimulus M was held constant with mean M ratings of 5.46 and 5.45 for the stimulus conditions of high and low I, respectively. No *S*s were dropped from the experiment.

TABLE 1

MEAN CORRECT ON TRIAL 1 AND MEAN CORRECT PER TRIAL DURING FREE RECALL OF PAIRED-ASSOCIATE (PA) RESPONSES IN EXPERIMENT I

| Free recall | PA trial | Imagery |  |  |  |
|---|---|---|---|---|---|
|  |  | Low |  | High |  |
|  |  | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* |
| Trial 1 | 3 | 3.37 | 1.89 | 3.73 | 1.61 |
|  | 4 | 3.45 | 1.82 | 5.81 | 1.10 |
| All trials | 3 | 5.33 | 1.94 | 6.20 | 1.10 |
|  | 4 | 5.55 | .88 | 7.02 | .57 |

### *Results*

The data for the learning of the PA lists were analyzed. The mean correct per trial for the groups with high stimulus I was 4.98, while the mean for the groups with low stimulus I was 2.95. This difference was significant, with $F\ (1,\ 40) = 14.62$, $p < .01$. No other difference was significant in the analysis. These data are presented in Table 1. In the analysis of the first trial of free recall (which came prior to the first observation trial of free recall), both imagery, $F\ (1,\ 40) = 8.18$, $p < .01$, and number of PA trials, $F\ (1,\ 40) = 5.32$, $p < .05$, were significant. In the analysis of the total number correct during free recall, the imagery variable was significant $F\ (1,\ 40) = 14.49$, $p < .01$. That is, trigrams previously associated with items such as TRIPOD and HURDLE were easier to recall than when they had been previously associated with such items as METHOD and THOUGHT. These results essentially replicate the results of Young and Parker (1970).

On the basis of the data of Exp. I, it would seem that the trigram responses employed in the PA lists acquired imagery, with greater degrees of imagery being acquired by responses associated with stimuli with high I than to stimuli with low I. Since the paradigm employed in Exp. I is essentially the same as that employed by Young and Parker (1970), it would appear that on the basis of these data, the acquisition of vividness, similarity, as well as a variety of other attributes of the verbal material itself could also be demonstrated. However, a more parsimonious explanation of the paradigm is that ease of stimulus recall is being manipulated. That is, it may be that a previously neutral response becomes easier to recall after being associated to an easily recalled stimulus not because the response has acquired meaningfulness, or imagery, etc., but because the easily recalled stimulus serves as a more efficient mediator for the recall of the response. If such is the case, then the acquisition of imagery, as in Exp. I, or meaningfulness, as in the Young and Parker study, has not been demonstrated; rather, the paradigm which has been used is one in which variable stimulus recall has been demonstrated. Experiment II was designed to investigate this possibility by using the same paradigm as that employed in Exp. I. In Exp. II ease of stimulus recall was manipulated, while at the same time stimulus attributes such as meaningfulness, imagery, etc., were held constant.

## EXPERIMENT II

### *Method*

*Design.*—Each of 90 *S*s learned a 16-pair PA list for eight trials and then learned the PA responses for five trials by free recall. Three groups were employed in the experiment: related cued (RC), related no cue (RN), and unrelated (U). The stimuli of the PA list learned by the related groups were composed of four categories with four words per category, and the stimuli of the U group were unrelated to one another. The RC group had the names of the stimulus categories posted on the wall during free recall of the responses and were told that this would help them "in some way" to recall the 16 trigrams. Recall for the RN and U groups was uncued. A common set of 16 low-M trigrams served as the responses for each group.

*Procedure and materials.*—The PA lists were presented for eight trials and each consisted of an observation and a test trial. Both the pairs and the stimuli were presented at a 5-sec. rate with a 20-sec. intertrial interval. Four orders of presentation were used during the observation and test trials to minimize serial position effects. Subsequent to the PA task, *S*s were given instructions for free recall and then were told to write down the previous PA responses. As in Exp. I, observation and test trials during free recall were alternated with a test trial coming first. During free recall, the items were presented at a 3-sec. rate. A 60-sec. interval was used during recall and the test trial—observation trial interval was 9-sec. Four orders of presentation were used.

The stimulus materials were taken from Battig and Montague (1969). The four most frequent responses in the body-parts, fruits, animals, and room-parts categories were used in the related lists, and a similarly frequent set of 16 items was taken from 16 different categories in the same norms for use in the unrelated lists. The trigram responses were taken from Archer's (1960) list and had a median association value of 20. The materials were presented on a Kodak 650 Carousel projector in a large room. The *S*s were run in groups of five each. Three *E*s were employed in the experiment and each ran two subgroups of *S*s in each of the three experimental conditions. The *S*s wrote their responses and used a card to cover the responses given during the same trial. After each trial, the answer sheet was placed face down on the table in front of *S*. Two *S*s were lost because of failure to follow instructions.

### *Results*

The number of correct responses given during Trials 7 and 8 during PA learning as analyzed. The per trial means for the U, RC, and RN groups were 15.47, 15.28, and 15.20, respectively. No main effects or interactions approached significance and hence the assumption of equivalent groups cannot be rejected. It should be noted that more responses were given correctly in the U groups than in the R groups. Although this difference was not significant, it replicated the results of Runquist (1970).

The number of correct responses given during free recall was also analyzed. The variability of the means of the three major groups was only marginally significant, $F\ (2, 72) = 2.84$, $p = .07$. However, since category versus no-category and cued versus uncued comparisons are implied by the design of the experiment, these orthogonal comparisons were made. The RC, RN versus U comparison was significant, $F\ (1, 72) = 5.49$, $p < .05$, but the two R groups did not differ between themselves with $F < 1.00$. It can be seen from Fig. 1 that the two R groups recalled more than the U group. Other significant differences list consisting of stimuli of either high or low I and a common set of low-M trigrams. Next, half of the *Ss* in each group were given four observation–test trials on an eight-pair A-D list consisting of the same stimuli as in the A-B list but with a different set of low-M trigram responses. The other half of the *Ss* in each group were given four trials on an eight-pair C-D list in which the stimuli were unrelated to the items of the first list but which had the same degree of I. The responses in the A-D and C-D lists were identical. Following second-list learning, the first-list stimuli (A) were presented at a 3-sec. rate and *Ss* were asked to give as many B responses as possible. All *Ss* then received four free recall trials in which only the first-list responses were presented on alternating observation and test trials with a test trial coming first. Finally, two *Es* were employed. Thus the design can be summarized as a 2 × 2 × 2 factorial with stimulus imagery (SI), paradigm (E vs. C), and *E* being factorially combined ($N = 10$ per cell).

*Procedure and materials.*—Lists were presented on a Kodak 650 Carousel projector. PA presentation was at a 3-sec. rate both for the stimuli alone (test trials) and for the pairs (observation trials). The intertrial interval was 6 sec. During free recall, the items were presented at a 3-sec. rate and the recall interval was 30 sec. The interval between test trial and observation trial was approximately 6 sec. The stimuli for the A-B, A-D lists had mean imagery ratings of 6.33 for the high-I and 2.70 for the low-I items (Paivio et al., 1968). Stimulus meaningfulness in the same lists was approximately equal with means of 5.50 and 5.45 for high- and low-I lists, respectively. The stimuli of the C-D lists had mean I ratings of 6.30 and 2.72 for the high and low groups, respectively, and had meaningfulness ratings of 5.71 and 5.51.. Responses were trigrams taken from Archer (1960) with mean association values of 21.2 and 20.7 for the B and D responses, respectively. Response similarity was low and no consonant was used more than once in a list.

Four different orders of presentation were used for the PA pairs of each list and for free recall to avoid possible serial learning. All responses were written in booklets with one page for each test trial. The *Ss* were run in a large room in groups of five each. No *Ss* were dropped from the experiment.

### *Results*

*First-list learning.*—The data from the learning of the first list were analyzed. The pairs of the low-SI condition were more difficult to learn than were the pairs of the high-SI condition, $F$ (1, 72) = 18.92, $p < .01$. These data are presented in Table 2.

*Second-list learning.*—Analysis of the second list indicates negative transfer occurred, with the performance of the C-D lists being greater than the A-D lists, $F$ (1, 72) = 6.54, $p < .05$. Care must be taken in the interpretation of these data. Different second lists were learned in the experimental and control conditions and transfer effects may be confounded with differential list difficulty. The high-SI conditions continued to perform at a higher level than did the low-SI conditions, $F$ (1, 72) = 9.60, $p < .01$. These data are also shown in Table 2.

TABLE 2

Mean Correct per Trial during First- and Second-List Learning during Experiment III

| List | SI | Paradigm | | | |
|---|---|---|---|---|---|
| | | A-D | | C-D | |
| | | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* |
| First | High | 3.23 | 1.57 | 2.78 | 1.34 |
| | Low | 1.66 | 1.35 | 1.80 | .88 |
| Second | High | 2.33 | 1.48 | 3.50 | 1.49 |
| | Low | 1.80 | 1.13 | 2.16 | 1.30 |

Note.—SI = Stimulus imagery.

*Retention.*—Following A-D learning, the A stimuli were presented and all *Ss* were tested for retention of the A-B associations of the first list. This was designed to determine if retention was differentially related to SI. It was found that retroactive inhibition (RI) had occurred, $F$ (1, 72) = 5.77, $p < .05$, and that recall in the high-SI groups was better than in the low-SI groups, $F$ (1, 72) = 13.30, $p < .01$. These data are presented in Table 3. No other *F*s were significant in the analysis, and it cannot be concluded that one group had greater amounts of RI than the other.

Subsequent to the test for retention of A-B associations, four free recall trials of the first-list responses were given. The analysis of these data indicated that first-list responses were better recalled after having been associated to stimuli with high imagery than to stimuli with low imagery, with $F$ (1, 72) = 6.70, $p < .05$. In addition, trials, $F$ (3, 216) = 53.83, $p < .001$, and Trials × Stimulus Imagery, $F$ (3, 216) = 2.81, $p < .05$, were significant. These

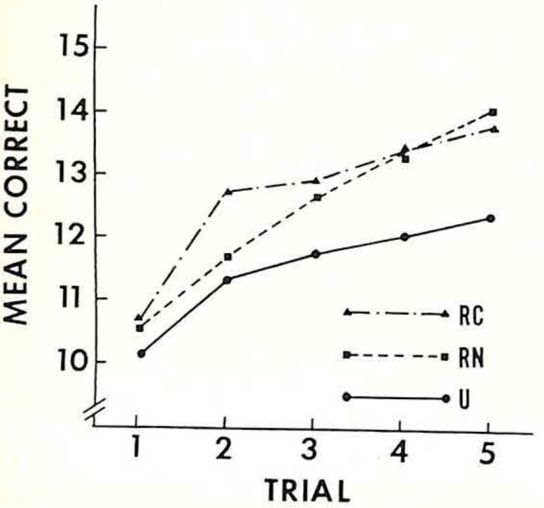


FIG. 1. Mean correct per trial during free recall of responses in Exp. II.

in the analysis were trials, $F$ (4, 288) = 53.12, $p < .001$, and the Linear Trend over Trials × Related versus Unrelated Groups interaction, $F$ (1, 288) = 6.98, $p < .01$. These data indicate that the R groups performance increased more rapidly over trials than did the U groups. No other comparisons approached significance in the analysis.

The modified ratio of repetition (MRR) reported by Bower, Lesgold, and Tieman (1969) was used to measure clustering during free recall. The 16 trigram responses used were unrelated, but recall of the trigrams may have been mediated by the stimuli to which they had become associated during PA learning. If such were the case, responses associated to stimuli belonging to the same concept should tend to be recalled together. As a control comparison, an MRR was computed for recall in the U group by assuming that the same trigrams belonged to the same stimulus items as those learned by the two R groups.

The MRR did vary across groups, $F$ (2, 72) = 25.57, $p < .001$. This variability was partitioned into RC, RN versus U and RC versus RN, comparisons. The former was significant, $F$ (1, 72) = 47.32, $p < .001$, while the latter barely failed to reach significance, $F$ (1, 72) = 3.82, $p$ = .06. The former comparison replicates Runquist (1970), and the latter indicates that clustering during cued recall tended to be greater than during uncued recall. Other significant differences in the analysis were trials, $F$ (4, 288) = 18.26, and the Trials × Groups interaction, $F$ (8, 288) = 4.16, both $p$s $< .001$. Virtually all of the variance associated with this latter comparison is accounted for by the Linear Trend over Trials × R versus U Groups interaction, $F$ (1, 288) = 32.05, $p < .001$. These data are presented in Fig. 2, and it can be seen that clustering as measured by MRR increased over trials for the R groups but did not for the U groups.

Data such as those presented in the recall and clustering analyses apparently argue strongly for the mediation hypothesis. However, it could also be argued that the trigram responses of the R groups acquired membership in their respective stimulus categories during PA learning and as such were clustered together and became easier to recall. Thus at least two hypotheses can account for the better recall of responses associated with categorized stimuli or with high-I stimuli: the mediation hypothesis and the acquisition hypothesis. Experiment III was designed to differentiate between these two hypotheses.

## EXPERIMENT III

### *Method*

*Design.*—Two groups of 40 *S*s each were given four observation-test trials on an eight-pair AB-

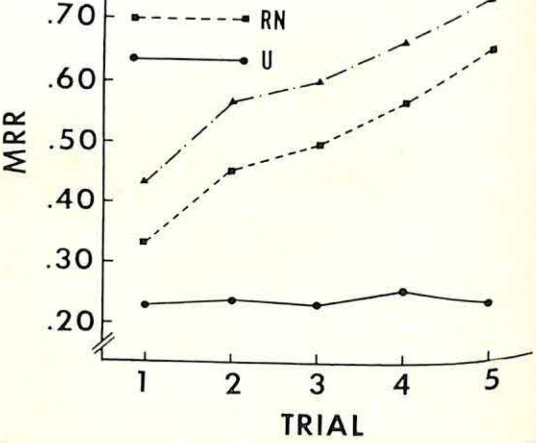


FIG. 2. MRR (clustering) scores during free recall of responses in Exp. II.

TABLE 3

Mean Numbers of Correct Responses Given during Test for Retroactive Inhibition (RI) and during the Four Trials of Free Recall of PA Responses in Experiment III

| SI | Cond. | RI | | Response recall trial | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | | 2 | | 3 | | 4 | |
| | | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* |
| High | A–D | 1.30 | 1.24 | 4.15 | 1.31 | 5.70 | 1.69 | 6.15 | 1.18 | 6.05 | 1.54 |
| | C–D | 2.40 | 2.01 | 4.15 | 1.79 | 5.50 | 1.57 | 6.35 | 1.23 | 6.70 | .92 |
| Low | A–D | .70 | .83 | 3.80 | 2.07 | 4.60 | 2.14 | 5.10 | 1.97 | 5.40 | 2.04 |
| | C–D | .95 | .68 | 4.10 | 1.89 | 4.75 | 1.94 | 5.40 | 2.31 | 5.80 | 1.54 |

Note.—SI = stimulus imagery.

data are also presented in Table 3. It is interesting to note that within conditions of SI, responses were equally available regardless of which second list was learned, $F(1, 72) = 1.28$, $p > .05$. That is, RI occurred during A-B recall but not during free recall of the B responses. No other *F*s were significant in the analysis.

## Discussion

Differential learning of responses previously learned in a PA task as a function of some characteristic of attribute of the PA stimuli has been fairly well demonstrated over a variety of conditions. Previous research has shown that the same set of trigrams is easier to learn in a serial list if the trigrams had previously served as responses in a PA list which contained high-M stimuli rather than low-M stimuli (Young & Parker, 1970). Experiment I extended this finding by demonstrating essentially the same results using I as a variable. A common set of trigrams is easier to learn in a free recall list if these trigrams have previously served as responses in a PA list containing high-I stimuli rather than in a list with low-I stimuli.

At least two hypotheses can explain the results of these experiments. The mediation hypothesis holds that responses associated to relatively easily recalled stimuli become themselves more easily recalled because the stimuli serve as mediators for response recall. An alternate hypothesis, called the acquisition hypothesis, holds that responses associated to relatively easily recalled stimuli acquire attributes of the stimuli which make the responses themselves easier to recall. Thus a neutral response associated to a stimulus with high-I comes to have an increased amount of I. A comparison of these two hypotheses was of especial interest in Exp. III.

In Exp. III, negative transfer was found during the learning of the second list and RI was found during A-B recall. Since no *E* versus C × SI interaction was found in any of these analyses, it can be concluded that amount of negative transfer and RI found was approximately the same over both conditions of SI. With RI occurring, it can be assumed that at least some of the forgetting observed during A-B recall was due to the unlearning of A-B associations in the Experimental groups. From the mediation hypothesis, this implies that free recall of first-list responses following A-B recall should be less than free recall in the comparable Control group. That is, according to the mediation hypothesis, RI should be demonstrated during response recall because the A-B link between the item to be recalled and its mediator was broken in the Experimental groups. RI was not demonstrated during response recall, and this fails to support the mediation hypothesis. It should be noted that in Exp. III, the test for free recall was made subsequent to the test for RI and that the free recall may have been influenced in some manner by the test for RI. If no test for RI had been made, however, there would have remained the possibility that no RI and no differential unlearning of A-B associations had occurred. Under these conditions, it would have been extremely difficult to argue that Exp. III represented a differential test of the mediation and acquisition hypotheses.

The response-recall data can provide more direct support for the acquisition hypothesis.

If unlearning of A-B associations reduces the ability of $S$ to recall responses using the appropriate stimuli as mediators, then (according to the mediation hypothesis) the difference between Experimental and Control groups with low-I stimuli should be less than the corresponding groups with high-I. That is, an $E$ versus C $\times$ SI interaction is predicted by the mediation hypothesis. On the other hand, if the response has acquired a certain degree of I (as per the acquisition hypothesis) during A-B learning, then unlearning of the A-B association should have relatively little effect upon response recall and the magnitude of the difference between Experimental groups with high and low SI should approximate the respective Control group difference. The response-recall data support the acquisition notion and give little, if any, support to the mediation hypothesis.

The conditions of Exp. II were replicated almost exactly by some of the groups in Runquist (1970). In contrast to Exp. II, Runquist failed to find that response recall was related to the class of stimulus material used. One difference between the two experiments which may have been of critical importance was response M. Whereas two-syllable adjectives were used as responses in the Runquist study, low-M trigrams were used in Exp. II. When both members of an S–R pair have approximately the same M value, the responses will be more easily recalled than will the stimuli (Asch & Ebenholtz, 1962). One implication of this is that $S$s might not use stimuli as mediators if the responses themselves are readily available. That is, support for the mediation hypothesis will be found only when the responses are more difficult to recall than the stimuli.

Although some support has been found for the mediation hypothesis, the data tend to give more support to the acquisition hypothesis. The data suggest that a previously neutral item is changed by virtue of having been associated with some other verbal unit. If the verbal unit has a considerable amount of I, then the previously neutral item acquires this imagery to some degree. However, there are other possible interpretations of the data and some of these will now be considered.

It may be that the conditions which showed greater free recall of previous PA responses did so as a result of differential familiarization during PA learning. Or it may be that greater numbers of correct responses during PA learning increased the availability of responses given in subsequent free recall (cf. Horowitz, Day, Light, & White, 1968). The trials variable of Exp. I was designed to be a familiarization control but as such failed to serve its purpose; the number of correct responses during PA learning in the high-I group with three trials was greater than in the low-I group with four trials instead of vice versa, as had been anticipated at the outset of the experiment. Thus in Exp. I, there is no control which eliminates familiarization as an explanation. However, the Young and Parker (1970) study contained some conditions with low stimulus M, wherein more PA responses were given correctly than in some conditions with high stimulus M. Subsequent learning of these responses in serial lists indicated the high-M conditions were learned more rapidly than the low-M conditions, apparently eliminating differential familiarization as an interpretation. Similarly, in Exp. II, more correct responses were given during PA learning to the U groups than to the R groups. Subsequent learning of the responses indicated the R groups were learned faster than the U group, again suggesting that familiarization is not a tenable explanation of the results of the experiment. Thus, while familiarization can be used on occasion as an alternate explanation, when the conditions of an experiment allow a comparison between familiarization and some other construct, little support is provided the former hypothesis.

The results of the present studies have implications concerning how verbal units are acquired outside the laboratory. If the results of the present series of studies are reliable, then the various attributes an item possesses may have been acquired through associations to other verbal units. Such associations could be acquired in a manner analogous to that employed in the present article or in a variety of other ways such as associating an item to the context which it is learned. Thus the kind and number of items to which a particular verbal unit is associated may provide the basis for determining the meaningfulness, imagery, etc., that the verbal unit possesses. These suggested methods of defining the degree of attributes possessed by a verbal unit are suggestive of procedures employed by Deese (1959) for determining strength of interitem associations and by Noble (1952) for determining degree of meaningfulness.

## REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, 74 (10, Whole No. 497).

Asch, S. E., & Ebenholtz, S. M. The principle of associative symmetry. *Proceedings of the American Philosophical Society*, 1962, **106**, 135–163.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 481–493.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, **5**, 305–312.

Horowitz, L. M., Day, R. S., Light, L. L., & White, M. A. Availability growth and latent verbal learning. *Journal of General Psychology*, 1968, **78**, 65–83.

Noble, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421–430.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, 241–263.

Paivio, A., & Madigan, S. A. Imagery and associative value in paired-associate learning. *Journal of Experimental Psychology*, 1968, **76**, 35–39.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76** (1, Pt. 2).

Runquist, P. A. Clustering in free recall following paired-associate learning. *Journal of Experimental Psychology*, 1970, **83**, 260–265.

Young, R. K., & Parker, G. V. C. Serial learning as a function of experimentally induced meaningfulness. *Journal of Experimental Psychology*, 1970, **84**, 24–26.

(Received August 20, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 151–155

# EFFECTS OF TEMPORAL AND SPATIAL ORGANIZATION OF LISTS ON CLUSTERING[1]

GAIL BRUDER[2] AND ERWIN SEGAL

*State University of New York at Buffalo*

Two studies investigated the effects of spatial as well as temporal organization on clustering and recall of categorized lists. In Exp. I, temporal blocking produced superior clustering compared with a random order of presentation. Spatial blocking, where items from the same category were presented in the same location, had no effect on performance. In Exp. II, when *S*s were informed of the nature of the list and of the arrangement of items, both spatial and temporal blocking augmented clustering. It was concluded that spatial location can be used by *S*s in organizing their recall.

While the phenomenon of clustering of related items during free recall has been well documented, agreement on the explanation for this has not been reached (see Tulving, 1968). Several investigators have suggested that associations or mediated associations are responsible for the tendency to recall related items together (Bousfield, Steward, & Cowan, 1964; Puff, 1966). Others have suggested that a strategy of coding, storing, and retrieving related items together is used by *S* (Cofer, Bruce, & Reicher, 1966; Cohen, 1966).

One of the more powerful variables influencing the degree of clustering has been the organization of the list during presentation. For example, when the related items are presented together (blocked), clustering is augmented relative to randomly ordered presentation conditions (Cofer et al., 1966). The effects of blocking can be handled by both explanations for clustering. Puff (1966) has argued that the priming of common mediating responses summates when related items appear in the same list, and that this summation is greater when the related items are temporally contiguous. On the other hand, blocked presentation could make the categorized nature of the list and the category names more accessible for use in coding. However, even when *S*s are given category names during presentation, blocked presentation still produces more clustering than random presentation (Dallett, 1964). This result was not conclusive since *S*-derived category labels may be more specific to the set of items.

The purpose of the present studies was to investigate whether spatial as well as temporal organization would influence clustering, and if so, to assess whether the results were more consonant with one of the explanations for clustering described above. Slamecka (1967), in investigating serial learning, found that serial recall of the spatial location of items was as good as serial recall of the temporal order. Slamecka separated temporal and spatial order by using a separate locus in a memory drum for each of the items. On the other hand, using a similar task, Mandler and Anderson (1971) found that serial recall of the temporal sequence was better than that of the spatial sequence whether or not *S*s were instructed as to the recall task before presentation of the lists. In the present studies, a similar technique was used to vary independently the temporal and spatial arrangement of related items and to investigate the effects on recall and clustering. Four conditions of list organization resulted from using either blocked or random arrangement of items for temporal order and spatial location. Since some conditions require more frequent shifts in attention to different windows, and this could influence learning, two control conditions were added. Items were arranged in either a random or blocked temporal order but the items always appeared in the same window location. Comparisons of the control and experimental conditions would indicate whether switches in window locations interfered with performance.

---

[1] The results of this research were reported at the Psychonomic Society meeting, St. Louis, November 1971. The authors thank Carol Dearlove for her assistance in collecting the data.

[2] Requests for reprints should be sent to Gail Bruder, Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Buffalo, New York 14226.

## EXPERIMENT I

### *Method*

*Design.*—A $3 \times 2 \times 3$ factorial design was employed. There were three variations in the assignment of items to window locations. In the control conditions, all words were presented in the center window of the memory drum (Sc); in the other two conditions, items were arranged spatially in a random fashion (Sr), or spatially blocked (Sb) according to category membership. For example in the Sb condition, all city names would appear in the same window location. The temporal order of presenting items was either randomly determined (Tr) or blocked by category membership (Tb). The six conditions were: Sc Tr, Sc Tb, Sr Tr, Sr Tb, Sb Tr, and Sb Tb, where the first two letters indicate the type of spatial arrangement of items and the last two letters, the type of temporal ordering of items. Recall and clustering performance was assessed over three trials.

*Materials.*—The items were selected by taking alternate items from the most frequent associates to category names (Battig & Montague, 1969). Ten items for each of 10 categories were obtained and two lists of 5 categories each were derived. Care was taken to avoid having similar categories or categories with common instances appear in the same list. For example, "CITIES" were used in one list and "COUNTRIES" in the other. Likewise "FRUITS" and "TREES" were assigned to different lists. The List 1 categories were: CITIES, FRUITS, ANIMALS, SPORTS, and PROFESSIONS. The List 2 categories were: COUNTRIES, TREES, VEGETABLES, BIRDS, and MUSICAL INSTRUMENTS. The spatial and temporal arrangement of items for presentation on the memory drum depended on the condition. Items were randomly arranged within those restrictions. The same relative order of items was maintained on the three trials, but the starting point was varied from trial to trial in order to minimize serial position effects on recall and recall order.

*Subjects.*—The *S*s were female introductory psychology students, who participated as part of a laboratory requirement. There were 10 *S*s in each of the six conditions. Half of the *S*s in each condition received List 1; the other half, List 2. Within these restrictions, a single table of random numbers was used to generate the order in which these conditions would be assigned. The *S*s were assigned to the conditions in the order in which they appeared in the laboratory.

*Apparatus.*—The lists were presented on a Lafayette memory drum with five window locations marked off by thin strips of black tape placed between locations.

*Procedure.*—Each *S* was seated in front of the memory drum and told that she would see 50 words presented 1 at a time in one of the windows of the memory drum. She was instructed to learn as many words as possible. She was further instructed that she would be free to recall the words in any order when the recall period occurred. The words were presented at a 2-sec. rate. Following the first presentation, *E* recorded the oral responses during a 3-min. recall period. Trials 2 and 3 proceeded in a similar manner. Following the third recall period, each *S* was questioned about the task and dismissed.

### *Results and Discussion*

The mean recall and clustering results over trials are indicated in Fig. 1. The results suggest that temporal, but not spatial, organization of related items influenced recall and clustering.

*Recall.*—The analysis of variance performed on the number of items recalled indicated that, except for a higher order interaction, only the effect of trials, $F\ (2, 96) = 331.99$, $p < .001$, was significant. The higher order interaction of Spatial Condition × Lists × Trials, $F\ (4, 96) = 5.11$, $p < .005$, was significant. Although we cannot explain this particular interaction, it does not affect any of the conclusions drawn from the study. The effect of temporal blocking approached significance, $F\ (1, 48) = 3.39$, $p < .10$. There was no effect of spatial condition ($F < 1$) nor did spatial condition interact with temporal condition ($F < 1$).

*Clustering.*—The Ratio of Repetition index ($\text{RR} = r/(N - 1)$) was used to assess clustering (Cohen, Sakoda, & Bousfield, 1954). The analysis of variance indicated that temporal blocking increased clustering, $F\ (1, 48) = 19.21$, $p < .001$. Clustering increased over trials, $F\ (2, 96) = 35.18$, $p < .001$. There was no significant effect of spatial condition, $F\ (2, 48) = 2.22$, $p < .25$. The Spatial × Temporal Conditions interaction was not significant, $F\ (2, 48) = 2.01$, $p < .25$. All other $F$s $< 1$.

Inspection of Fig. 1 suggests there is no simple effect on performance of requiring shifts in attention to different window locations during list presentation. When

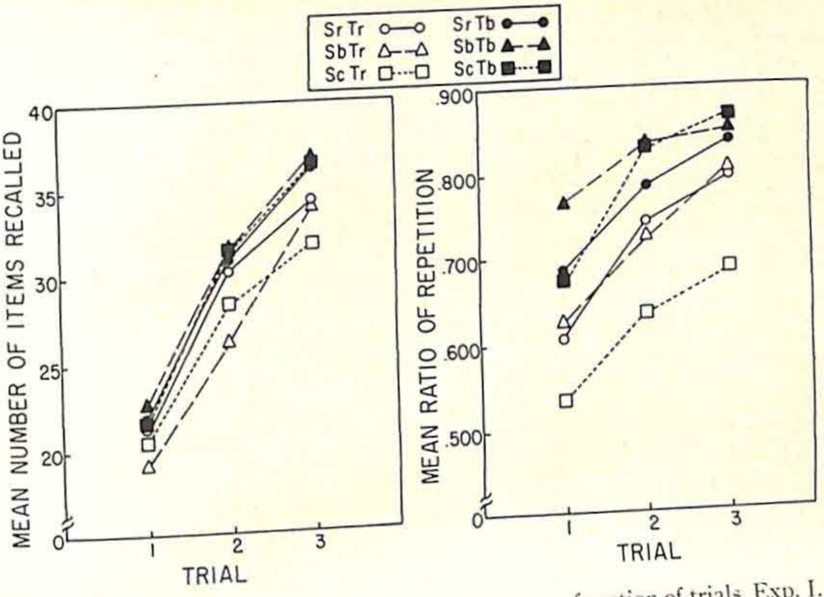


Fig. 1. Mean recall and ratio of repetition scores as a function of trials, Exp. I.

items are temporally blocked, constant shifting of windows seems to produce inferior clustering (compare Sr Tb with Sc Tb and Sb Tb conditions). However, when items are randomly ordered (temporally), the shifting of locations seems to augment clustering (compare Sb Tr and Sr Tr with Sc Tb). It is tempting to speculate that shifts of attention from one spatial location to another may interrupt the establishment of rehearsal sets based on temporal contiguity. If items in the same rehearsal set tend to be recalled together, interruption of rehearsal of contiguous items could influence recall order and therefore, clustering. Clustering would be reduced by this interruption if temporally contiguous items were related, (Sr Tb condition) and augmented by an interruption if temporally contiguous items were not related (Sr Tr and Sb Tr conditions). However, this hypothesis would require additional testing not provided by this study.

Since only temporal organization influenced recall and clustering, the results tend to support Puff's (1966) hypothesis that the summation of priming effects and therefore, clustering, increases with temporal contiguity of the related items.

However, it seemed possible that the differences between temporal and spatial bases of organization might lie in the ease of discovering the organization. While temporal blocking can be discovered by comparing successive items, discovery of spatial blocking requires remembering the window location of the previous instance of the category. Therefore, the experiment was replicated with instructions added in order to make the spatial basis of organization more obvious to *S*s.

## Experiment II

Experiment II replicated Exp. I with two modifications. First, the control conditions were eliminated, since the main concern was with further testing the effectiveness of spatial organization on performance. The second modification was instructing *S*s about the categorical nature of the list and how the list was organized. The materials and the apparatus were the same as in Exp. I.

*Design.*—The design was basically the same with only four of the conditions, namely: Sr Tr, Sb Tr, Sr Tb, and Sb Tb.

*Subjects.*—The *S*s were 56 female introductory psychology students. The *S*s were randomly assigned to conditions and lists as they arrived at the laboratory.

*Procedure.*—The procedure was the same as in Exp. I, except for the additional instructions. All *S*s were told that the

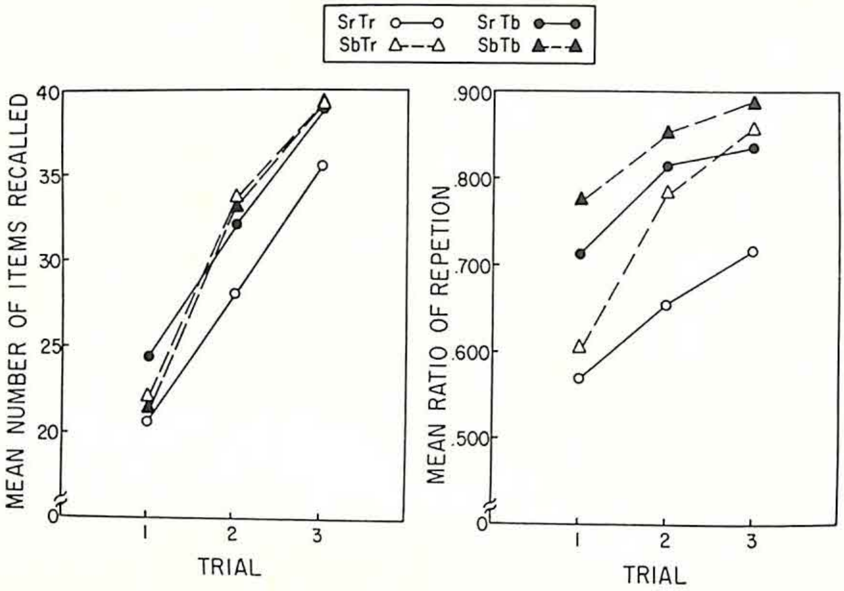


FIG. 2. Mean recall and ratio of repetition scores as a function of trials, Exp. II.

list contained words belonging to five categories. In addition, *S*s in the conditions with temporal blocking were told that words from the same category would always be shown consecutively. The *S*s in the conditions with spatial blocking were told that words from the same category would always appear in the same window location and that there were five window locations. The *S*s in Cond. Sb Tb received instructions about both temporal and spatial organization, in that order.

### *Results and Discussion*

The mean recall and clustering scores over trials are shown in Fig. 2. The organization of the lists affected both the number of words recalled and clustering.

*Recall.*—The main effect of trials was significant, $F$ (2, 96) = 295.27, $p < .001$. Neither spatial condition, temporal condition, nor the Spatial × Temporal interaction reached significance in recall, $F$s (1, 48) = 1.25, $p > .25$, and 1.37, $p < .25$, and 2.07, $p < .25$, respectively. However, all three blocked presentation groups showed a higher performance than the nonblocked groups, and a blocked–nonblocked linear comparison reached significance, $F$ (1, 48) = 4.63, $p < .05$. The Spatial Condition × Trials interaction also reached significance, $F$ (2, 96) = 4.74, $p < .05$. Thus blocking of the input augmented the number of words recalled by a small amount. After the first trial, any organized input had equal effect on recall.

*Clustering.*—The effect of trials was again significant, $F$ (2, 96) = 59.40, $p < .001$. Also, under these conditions, *S*s were able to use the consistent spatial locations of category items to facilitate recalling related items together. Both spatial and temporal blocking augmented clustering, $F$s (1, 48) = 7.48, $p < .005$, and 18.02, $p < .001$, respectively. The Temporal Condition × Trials interaction was significant, $F$ (2.96) = 3.69, $p < .05$. Separate analyses of each trial indicated that while statistically significant on all trials, the effect of temporal blocking diminished over trials, $F$s (1, 48) = 19.16, 11.34, and 8.39 for Trials 1, 2, and 3, respectively. Spatial blocking showed an increasing effect, and was significant only on Trials 2 and 3, $F$s (1, 48) = 5.89, $p < .05$, and 12.94, $p < .001$, respectively; for Trial 1, $F$ (1, 48) = 1.74, $p < .25$.

One possible explanation for better performance in the blocked groups in this experiment is that the extra information in the instructions led *S*s in these groups to more easily identify the categories used by *E*. If *S*s in the random group (Sr Tr), more than *S*s in the other groups, categorized the items in the list according to *S* units that did not correspond to *E* units, the results could be deemed artifactual. This possibility was minimized by selecting relatively high-frequency members of disparate categories in constructing a stimulus list. However, to check the possibility of misclassification, extralist intrusions in the Sr Tr condition were examined.[3] There were a total of 15 extralist intrusions. Thirteen were members of *E* categories. The other 2 were contributed by one *S* who twice gave "MARYLAND" from the list containing cities, one of which was BALTIMORE. Thus, the results seem to be a function of instructions making the organization of the lists evident and available for use by *S*s. The data from the postexperimental questions provide additional support for this conclusion. The *S*s were asked whether the window location of the words helped them recall the words. In Exp. I, half of the *S*s in the Sb Tb and Sb Tr conditions said "no." In Exp. II, none of the *S*s in the Sb Tr condition said "no." However, slightly fewer than half of the *S*s in the Sb Tb condition said "no." This difference in the Sb Tb and Sb Tr conditions suggests *S*s were not just replying "yes" because spatial location was mentioned in the instructions. In addition, it suggests that when given two bases of organization, temporal and spatial (Cond. Sb Tb), some *S*s may select which to use.

## General Discussion

The focus of these experiments was the effect of spatial blocking on the recall and organization of categorized word lists. It is evident that spatial blocking per se had no effect on clustering or recall. It is only after *S*s were given instructions that each window in the memory drum was associated with a particular category of items that spatial blocking had an effect. After two trials under these conditions, spatial blocking was as good as temporal. It may be that *S*s have to learn the category association to the windows before its spatial location aids clustering.

These results strain a simple temporal contiguity explanation of augmented clustering as a function of blocked input. Even though *S*s did not always show greater clustering as a function of spatial organization, they did so with instructions. This suggests that spatial coding of the categorized items is at least a two-stage process. The *S*s must first learn which categories are associated with which locations before they use the locations to organize storage and/or retrieval. The uninstructed groups never learned the first stage. It seems that this difference between the two experiments is due to process differences as a function of instructions rather than directly to associative differences. The associative priming mechanism (Puff, 1966) can be modified to allow for implicit elicitation of category names as a function of the spatial arrangement and instructions, but then this theory and the strategy theory (Cohen, 1966) are no longer incompatible.

## REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Bousfield, W. A., Steward, J. R., & Cowan, T. M. The use of free associational norms for the predicting of clustering. *Journal of General Psychology*, 1964, **70**, 205–214.

Cofer, C. N., Bruce, D. R., & Reicher, G. M. Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 1966, **71**, 858–866.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 182–187.

Cohen, B. H., Sakoda, J. M., & Bousfield, W. A. The statistical analyses of the incidence of clustering in the recall of randomly arranged associates. (ONR Tech. Rep. No. 10, Contract NOMR-631(00)) Washington, D. C.: United States Government Printing Office, 1954.

Dallett, K. M. Number of categories and category information in free recall. *Journal of Experimental Psychology*, 1964, **68**, 1–12.

Mandler, G., & Anderson, R. E. Temporal and spatial cues in seriation. *Journal of Experimental Psychology*, 1971, **90**, 128–135.

Puff, C. R. Clustering as a function of the sequential organization of stimulus word lists. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 503–506.

Slamecka, N. J. Serial learning and order information. *Journal of Experimental Psychology*, 1967, **74**, 62–66.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

(Received August 21, 1971)

---

[3] The authors wish to thank Leo Postman for suggesting that the extra list intrusions be examined.

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 156–162

# EFFECT OF REPETITION OF STANDARD AND COMPARISON TONES ON RECOGNITION MEMORY FOR PITCH[1]

DIANA DEUTSCH[2]

*Center for Human Information Processing, University of California, San Diego*

In all conditions, pitch recognition was required after a 4-sec. interval during which four other tones were played. The effects of including in the interpolated sequence a tone at the same pitch as the standard (S) or comparison (C) tone were investigated. When the S and C tones were identical in pitch, inserting in the interpolated sequence a tone at that pitch caused a reduction in errors. When the S and C tones differed in pitch, inserting a tone at the pitch of the S tone produced a reduction in errors and inserting a tone at the pitch of the C tone produced an increase in errors. Inserting one tone at the pitch of the S tone and another at the pitch of the C tone produced no significant change in errors, compared with the condition in which no tone at either pitch was included in the interpolated sequence.

Several studies have been made of recognition memory for pitch when the tones to be compared are separated by a retention interval. Koester (1945) has demonstrated that memory for both loudness and pitch deteriorates with time, using the differential threshold (DL) as a measure of memory. Bachem (1954) found a continuing deterioration of pitch recognition up to a week, yet showing that we can maintain some sort of memory for an auditory stimulus over very long periods. Harris (1953) measured the DL with various interstimulus intervals (ISIs) under two conditions. When the standard tone was fixed at 1,000 cps, the DL was found to remain constant at around 4 cps for ISIs of up to 3.5 sec., with a decline of only .8 cps with an ISI of 15 sec. With a roving standard tone varying in frequency from 950 to 1050 cps, no increase in the DL occurred up to a 1.0-sec. ISI. Then a linear increase in the DL was found up to 15-sec. ISI where the DL was increased by 3.7 cps. It can be concluded that memory for pitch decays spontaneously with time. However, the amount of deterioration manifest in 15 sec. is small.

Until recently, the effects of interpolated stimuli on pitch recognition have not been investigated. Wickelgren (1966, 1969) has found that a tone incorporated between the standard (S) and comparison (C) tones produces memory deterioration which increases with increased duration of the intervening tone. In an experiment designed to study the effects of increased interference in a retention interval of constant duration, the present author presented *S*s with a set of S and C tones, each separated by a $4\frac{1}{4}$-sec. retention interval. A sequence of tones was interpolated during this retention interval, always beginning at 300 msec. after the end of the S tone and ending 2 sec. before the start of the C tone. The interpolated tonal sequence consisted of either four, six, or eight tones spaced regularly within the remaining time span. A statistically significant increment in errors was obtained, both when the number of tones interpolated during the retention interval was increased from four to six and also when it was increased from six to eight. It can be concluded that memory for tonal pitch is subject to interference by other tones even when the retention interval is held constant. Massaro (1970) has also reported that increasing the number of tones interpolated during an interval of constant duration increases the number of errors in pitch recognition. In this experiment, Massaro varied the number of interpolated tones between one and four.

---

[1] The work reported in this paper was performed in partial fulfillment of the requirement for a PhD thesis at the Department of Psychology, University of California at San Diego. I thank my advisor, G. Mandler, and the members of my thesis committee, A. Garsia, D. M. Green, I. Jacobs, and D. E. Rumelhart, for their advice and encouragement. I also thank D. A. Norman for use of computer facilities supported by United States Public Health Service Training Grant NS-07454-06. The work was supported in part by United States Public Health Service Training Grant MH-10835 and Grant MH-15828-01.

[2] Requests for reprints should be sent to Diana Deutsch, Department of Psychology, P.O. Box 109, University of California, San Diego, La Jolla, California 92037.

The above findings show clearly that memory for pitch does not simply deteriorate with time but is subject to some interference effect. A further experiment (Deutsch, 1970b) has shown that this interference cannot be explained in such general terms as a distraction of attention or a limitation in information storage capacity. In this study it was shown that the incorporation during the retention interval of spoken numbers, which *Ss* are required later to recall, produces only a minimal decrement in the same pitch-recognition task as is severely disrupted by the interpolation of other tones. Interactive effects must therefore take place within the pitch memory store itself. The following experiment represents a further investigation into the pitch memory store. It was theorized that there must be a representation in memory both of the item to be remembered, and also of the time or order in which this item occurred. One can obtain information about these representations by repeating the same item at a different point in time, since here item and order information are not simultaneously covaried. This procedure was used in the experiment described here. A preliminary investigation of this nature is reported in Deutsch (1970a).

## METHOD

*Procedure.*—In all experimental conditions, *Ss* listened to an S tone which was followed by a sequence of four interpolated tones, and later, after a pause, by a C tone. The *Ss* were instructed to try to remember the S tone, ignore the four interpolated tones if they wished, and then to judge whether the C tone was or was not the same in pitch as the S tone. They then indicated their judgments by writing "same" or "different" on paper. All the tones were of equal loudness, and 200 msec. in duration. The interval between the S tone and the first interpolated tone was 300 msec. and the interpolated tones were also spaced 300 msec. apart. A 2-sec. pause was incorporated between the last interpolated tone and the C tone.

*Conditions.*—The different conditions of the experiment are shown in Fig. 1. It can be seen that the presence or absence in the interpolated sequence of a tone at the pitch of the S tone was systematically varied; as was the presence or absence of a tone at the pitch of the C tone. Further, the placement of these critical tones in the interpolated sequence was systematically varied between the second and third serial positions. The entire tape consisted of 96 sequences. These were presented in 12 groups of eight. Sequences within each group of eight were separated by 10-sec. pauses; and there were 2-min. pauses between the groups. The sequences were arranged in random order with no separation by condition except that each group of eight contained four sequences in which the S and C tones were the same in pitch, and four in which they were different, to minimize the development of response biases. Also, no group of sequences contained more than one example of each S and C tone pitch (see below) so that pitch repetition effects for these critical tones were minimized. The *Ss* listened to the entire tape on two different occasions and the results were averaged.

*S and C tones.*—S and C tone pitches were taken from an equal-tempered scale (international pitch: A = 435) ranging from the C♯ a semitone above middle C to the C an octave above. The frequencies employed were: C♯ = 274, D = 290, D♯ = 308, E = 326, F = 345, F♯ = 366, G = 388, G♯ = 411, A = 435, A♯ = 461, B = 488, C = 517. Each of these tonal pitches was employed equally often within each condition, either as an S tone, or as a C tone, or both. When the S and C tones differed in pitch, in half of the sequences the S tone was higher than the C tone and in the other half it was lower.

*Interpolated tones.*—The interpolated tonal pitches were also taken from the equal-tempered scale (A = 435) and ranged from middle C to the C♯ an octave above. The frequencies employed were: C = 259, C♯ = 274, D = 290, D♯ = 308, E = 326, F = 345, F♯ = 366, G = 388, G♯ = 411, A = 435, A♯ = 461, B = 488, C = 517, C♯ = 561. The tones in the intervening sequences were chosen randomly from this set, except that no sequence contained repeated tones unless specified by the experimental condition. This restriction also covered repetition of tones separated by exactly an octave.

*Subjects.*—Fifteen *Ss* participated in this experiment. These were 14 University of California at San Diego undergraduates and 1 housewife and they were paid for their participation. The *Ss* were selected on the basis of obtaining a score of at least 75% correct on a small tape containing similar sequences. About one out of five applicants was accepted by this procedure.

*Apparatus.*—The tones were generated by a Wavetek oscillator controlled by a PDP-9 computer. The output was recorded on high-fidelity tape. The tape was played to *Ss* on a high-quality tape recorder through loudspeakers with the loudness at a comfortable listening level.

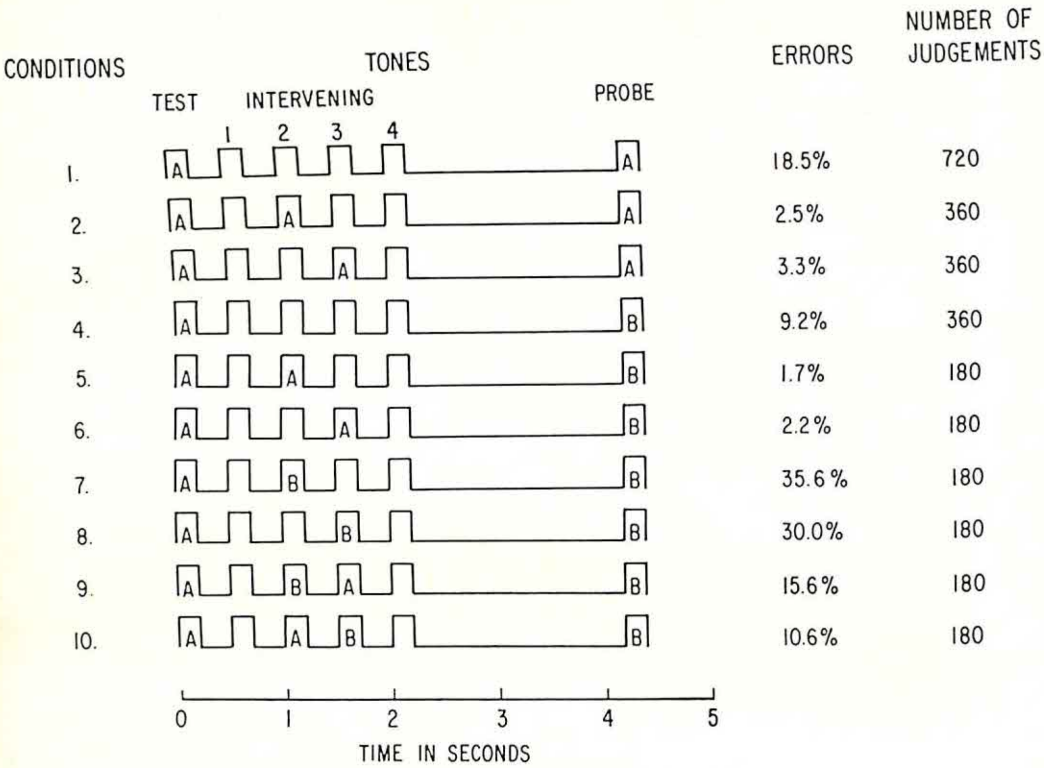


FIG. 1. The different conditions of the experiment with number of judgments and percentage of errors made in each. (Each condition is diagramed by a set of rectangular elevations, each of which indicates the presentation of a tone. The scale at the base of the figure indicates the time during which each tonal sequence is played. The initial elevation in each sequence indicates that the S tone occurred. The next four elevations labeled 1, 2, 3, and 4 indicate that the interpolated tones occurred, and show their serial position. The final elevation indicates that the C tone occurred. The initial elevations in all the sequences are labeled A and indicate the S tone. When the symbol A occurs elsewhere in a sequence, this indicates that a tone at the same pitch as the S tone occurred at that time. Those final elevations which are not labeled A are labeled B. This indicates that the C tone differed in pitch from the S tone. An elevation in the intervening sequence labeled B indicates that a tone at the pitch of the C tone (when this differed from the S tone) occurred in the indicated position of the intervening sequence.)

## RESULTS

The results of this experiment are shown in Fig. 1.

### *Effects of the Interpolated Tones*

The effect of inserting a tone at the pitch of the S tone, when the S and C tones are identical in pitch is shown by comparisons of Cond. 1 with Cond. 2 and 3. A considerable reduction in errors is produced. For both comparisons, this improvement is significant on sign tests beyond the .001 level.

The effect of inserting a tone at the pitch of the S tone when the S and C tones differ in pitch is shown by comparisons of Cond. 4 with Cond. 5 and 6. An improvement in detection of a difference between the S and C tones is produced. This effect is significant for both comparisons on Wilcoxon tests beyond the .01 level (two-tailed).

The effect of inserting a tone at the pitch of the C tone, when the S and C tones differ in pitch is shown by comparisons of Cond. 4 with Cond. 7 and 8. It can be seen that a striking increase in errors occur. For both comparisons, this effect is significant on sign tests beyond the .001 level.

The effect of inserting two tones, one at the pitch of the S tone and the other at the pitch of the C tone when the S and C tones differ in pitch, is shown by comparison of Cond. 4 with Cond. 9 and 10. Here there are no significant differences in the number of errors, on Wilcoxon tests ($p > .05$, two-tailed). Thus, the dramatic increase in errors produced when the C-tone pitch alone is included in the intervening sequence does not persist when the S-tone pitch is also included in the same sequence. In fact, the reduction in errors produced both when Cond. 9 is compared with Cond. 7, and also when Cond. 10 is compared with Cond. 8 is significant on Wilcoxon tests beyond the .01 level (two-tailed). Thus the benefit to discrimination produced by inserting the S-tone pitch in the intervening sequence effectively counteracts the damage produced by inserting the C-tone pitch in the same sequence.

### *Serial Position Effects*

Including in the intervening sequence a tone at the pitch of the S tone produced a greater reduction in errors when the critical tone was in the second serial position than in the third. This was true both when the S and C tones were identical (Cond. 2 compared with Cond. 3) and also when the S and C tones differed in pitch (Cond. 5 compared with Cond. 6).

Including in the intervening sequence a tone at the pitch of the C tone, when the S and C tones differed in pitch, produced a greater increase in errors when the critical tone was included in the second serial position than in the third (Cond. 7 compared with Cond. 8).

When tones at both the pitches of the S tone and the C tone were included in the same intervening sequence, errors were greater when a tone at the pitch of the C tone was included in the second serial position and a tone at the pitch of the S tone in the third, than when the serial positions of these critical tones were reversed (Cond. 9 compared with Cond. 10). This difference is significant on a Wilcoxon test at the .05 level (two-tailed).

Although only one of these effects reached statistical significance, all the trends described are consistent with the following generalization: When the included tone was closer in serial position to the S tone, there was an increased tendency to identify its pitch as that of the S tone.

### *Effect of Lower Compared with Higher S- and C-Tone Pitches*

With the S- and C-tone pitches in the lower half of the range employed in this study (274–366 cps), errors were greater than with the S- and C-tone pitches in the upper half of the range (388–517 cps). Errors in these two conditions were at 15.5% and 9.7%, respectively. This effect is significant on a Wilcoxon test beyond the .01 level (two-tailed). It may be explained by considering that the equal-tempered scale is a ratio scale, and so absolute differences in terms of cps increase as the scale is ascended; yet the DL for pitch over the range employed here remains roughly constant in absolute terms as the scale is ascended (Shower & Biddulph, 1931). Discrimination would therefore be easier at the higher frequencies since here the absolute-pitch differences between S and C tones are greater. In this experiment, however, all S- and C-tone pitches occurred equally often in all conditions.

## DISCUSSION

An attempt at rigorous theoretical quantification in explaining these results appears premature; however, a theory can be suggested which seems to be most in accord with the findings.

It is proposed that memory for the pitch of a tone is laid down simultaneously both on a pitch continuum and also on a temporal continuum. This may be represented as in the three-dimensional diagram in Fig. 2. A tone is here represented as laid down in memory in the form of a three-dimensional bell-shaped distribution. As time proceeds, this distribution spreads in both directions, but particularly along the temporal continuum. Also with the passage of time, each distribution shrinks in height, and so the more recent the tone the larger the distribution representing it. When the distributions underlying two tones overlap, the overlapping portions sum. In this way, summation between distributions underlying tones presented at different points in time will occur, due to the spread of these distributions along the temporal continuum.

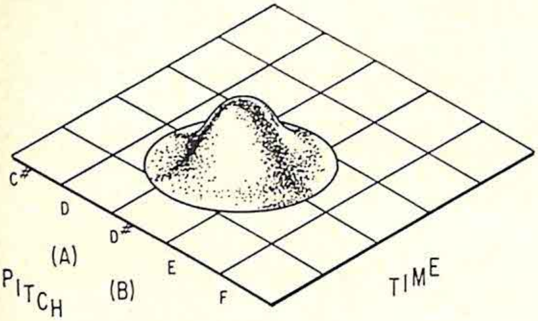


FIG. 2. Proposed distribution underlying memory for the pitch of a tone.

It is further proposed that the decision as to whether a given tone presented at one point in time (in this experiment, the C tone) is or is not the same in pitch as another tone presented at a different point in time (in this experiment, the S tone) is made in the following way. Assume that *S* divides the pitch continuum arbitrarily into sections according to his knowledge of the quantal steps employed in the experiment. He similarly divides the temporal continuum into sections. He is also able to estimate the volume of distribution within each three-dimensional section (see Fig. 2). Now when the C tone sounds, *S* examines the distribution in the section which corresponds to the same pitch as that of the C tone, and to the same temporal region as that of the S tone. He then compares the volume of distribution in this section with those in sections which are adjacent along the pitch continuum, but in the identical temporal region. If the volume of distribution in this section is much larger than those in neighboring sections in the same temporal region, *S* concludes that the S and C tones were the same in pitch. Conversely, if the volume of distribution in this section is much smaller than in a neighboring section in the same temporal region, *S* concludes that the S and C tones were different in pitch. As the volume of distribution in the appropriate section becomes more similar to that in a neighboring section in the same temporal region, the decision as to whether the S or C tones were the same or different in pitch becomes increasingly more difficult and so errors more numerous. Thus, the degree of accuracy of discrimination is determined by the size of a ratio between two distribution volumes.

The diagrams in Fig. 3 represent cross sections of the model on Fig. 2 taken at the point along the temporal continuum representing the presentation of the S tone. The two adjacent sections are arbitrarily labeled A and B to conform to the terminology in Fig. 1 and represent a semitone difference in pitch. The S tone is always A and the C tone is either A or B.

In Fig. 3, a, no tone either at pitch A or at pitch B occurs in the intervening sequence. The distribution at this point in time is therefore almost entirely due to presentation of the S tone A. (There will be some spread due to presentation of other tones at least a whole tone removed from A but for the purpose of simplicity this is not considered at present: In this experiment, such tones are randomly presented.) There is therefore a substantial difference in the amount of distribution in the A section compared with the B section; however, errors will still be made due to the distribution in the B section.

In Fig. 3, b, a tone at the pitch of the S tone (A) is included in the intervening sequence. The distribution underlying this included tone will spread along the temporal continuum and sum with that already existing due to presentation of the S tone A. The result shown in Fig. 3, b, will be an increased volume of distribution most specifically in the A section of the pitch continuum. The proportional difference between the volume of distribution in the A section and neighboring sections will therefore be increased; and discrimination will become more accurate. This would explain the decrease in errors produced by inclusion of the S-tone pitch in the intervening sequence, both when the S and C tones are identical in pitch (Cond. 2 and 3 compared with Cond. 1) and also when the S and C tones are different in pitch (Cond. 5 and 6 compared with Cond. 4). Further, the closer in time this included tone is to the S tone, the more will its distribution overlap with that of the S tone and the greater will be the resultant improvement in discrimination. This would explain the greater improvement in performance found when the included tone A is presented in the second serial position of the intervening sequence (Cond. 2 and 5) compared with the third serial position (Cond. 3 and 6).

In Fig. 3, c, a tone at the pitch of the C tone B is included in the intervening sequence. The distribution representing this included tone will spread along the temporal continuum and sum with that already existing due to presentation of the S tone A. The distribution existing at the point along the temporal continuum representing the presentation of the S tone A will therefore look somewhat as in Fig. 3, c. It is can be seen here that the proportional difference between the volume of

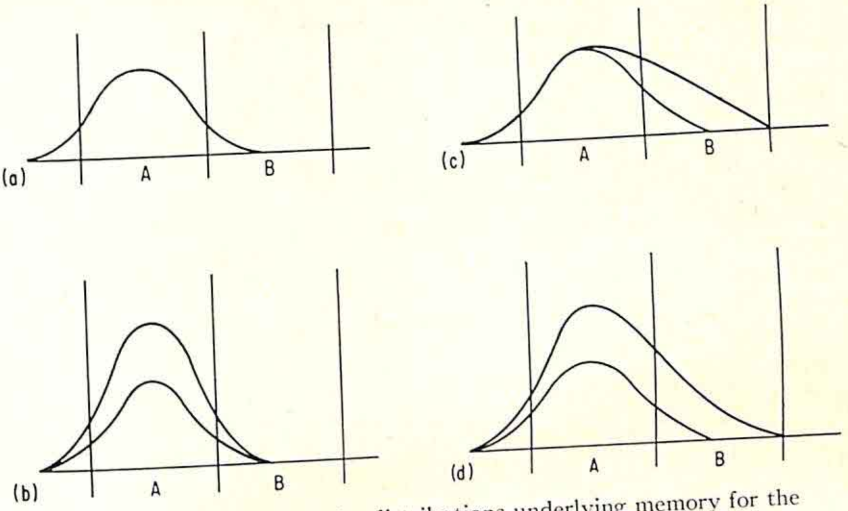


FIG. 3. Cross sections of the distributions underlying memory for the pitch of a tone under different experimental conditions.

distribution in Sections A and B is now decreased; so discrimination would become more difficult and errors more numerous. This would explain the increase in errors in Cond. 7 and 8 compared with Cond. 4. Further, the closer in time the included tone B is to the S tone A, the more will its distribution overlap with that of the S tone, and so the greater will be the decrement in performance. This would explain the greater increase in errors found when the included tone B is presented in the second serial position of the intervening sequence (Cond. 7) compared with the third serial position (Cond. 8). One should also expect from this theory that accuracy of performance in Cond. 1 (S and C tones identical in pitch, with this pitch not included in the intervening sequence) should depend very much on whether or not a tone a semitone removed from the S tone is included in the intervening sequence. Such an inclusion should produce a substantial increase in errors. Although there is no condition in this experiment which formally tests this prediction, it was found that in 8 of the sequences in Cond. 1, there was no such inclusion; and in 16 of the sequences, a tone a semitone removed from the S tone did indeed occur in the intervening sequence. Analysis of errors on this basis showed that where there was no such inclusion, errors were only 12.1% and where there was such an inclusion errors increased to 21.7%. Such a finding would be expected on the present theory.

In Fig. 3, d, tones both at the pitch of the S tone and also at the pitch of the C tone are included in the intervening sequence. This causes both the effects shown in 3, b, and 3, c, to occur. Thus, the volumes of distribution in both Sections A and B are increased; and so a considerable difference between them is retained. Discrimination should therefore be relatively easy. This would account for the nullification of the damage to discrimination produced by inclusion of a tone at the pitch of the C tone in the intervening sequence (Cond. 7 and 8) by the inclusion also of a tone at the pitch of the S tone (Cond. 9 and 10). Further, a powerful serial position effect should be expected in this condition, since here the serial positions of the included tones at both Pitch A and at Pitch B are simultaneously reversed. And a statistically significant serial position effect is indeed found when Cond. 9 and 10 are compared. One should also expect from this theory that accuracy of performance in those sequences in which the S and C tones were the same in pitch, with a tone at that pitch included in the intervening sequence (Cond. 2 and 3), should depend on whether or not a tone a semitone removed from the S-tone pitch is also included in the intervening sequence. Although there is no condition which formally tests this prediction it was found that in 15 out of the 24 sequences in Cond. 2 and 3 there was no such inclusion and in 9 of the sequences such an inclusion did occur. Analysis of errors on this basis showed that where there was no such inclusion, errors were 2.4%, and where such a tone was included, errors rose to 3.7%. This increment in errors was in the direction predicted by the theory.

It is apparent from this study, as well as from that described in Deutsch (1970b), that tonal pitch deteriorates rapidly in the presence of other tones. This finding has important bearing on theories of musical information storage. It has been suggested by several theorists (Cohen, 1962) that musical information (with particular reference to pitch information) is retained in its absolute form in an information store of limited channel capacity. However, the rapid deterioration of pitch information found in these experiments demonstrates that most of us are incapable of short-term storage of more than a minimal amount of absolute-pitch information, in a situation where other tones are also presented. Further, even when highly selected *S*s are used, so that the average level of performance on this task is very high, it is demonstrated in the present experiment that the presence of a critical tone in the middle of a sequence can have a highly deleterious effect on the recognition of the first tone in that sequence. We may recognize that a certain tone had occurred in a sequence without being able to tell whether it was or was not the first tone in that sequence.

It must be concluded from these findings that we cannot store transitional probabilities between tones with any success. It is suggested instead that information concerning absolute pitch is rapidly discarded, and that we store pitch information in an abstracted or recoded form. This conclusion is quite consistent with everyday knowledge of musical information processing. We recognize a tune much more readily than we recognize what key it was played in. The same argument holds for recognition of harmonic sequences; since not only are these transposable, but also their component chords are invertable. A theory of how pitch information is abstracted has been described by Deutsch (1969).

## REFERENCES

BACHEM, A. Time factors in relative and absolute pitch determination. *Journal of the Acoustical Society of America*, 1954, **26**, 751–3.

COHEN, J. E. Information theory and music. *Behavioral Science*, 1962, **7**, 137–163.

DEUTSCH, D. Music recognition. *Psychological Review*, 1969, **76**, 300–307.

DEUTSCH, D. Dislocation of tones in a musical sequence: A memory illusion. *Nature*, 1970, No. 5242, 226. (a)

DEUTSCH, D. Tones and numbers: Specificity of interference in short term memory. *Science*, 1970, **168**, 1604–1605. (b)

HARRIS, J. D. The decline of pitch discrimination with time. *Journal of Experimental Psychology*, 1952, **43**, 96–99.

KOESTER, T. The time error in pitch and loudness discrimination as a function of time interval and stimulus level. *Archives of Psychology, New York*, 1945, No. 297.

MASSARO, D. W. Retroactive interference in short-term recognition memory of pitch. *Journal of Experimental Psychology*, 1970, **83**, 32–39.

SHOWER, E. G., & BIDDULPH, R. Differential pitch sensitivity of the ear. *Journal of the Acoustical Society of America*, 1931, **3**, 275–287.

WICKELGREN, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1966, **72**, 250–259.

WICKELGREN, W. A. Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, 1969, **6**, 13–61.

(Received August 23, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 163–168

# SPATIAL S–R COMPATIBILITY EFFECTS INVOLVING KINESTHETIC CUES

RICHARD J. WALLACE[1]

*University of Oregon*

The *S*s were required to press either of two keys placed to their left and right when one of two stimuli (either an outline circle or square) was presented. The figure could appear in one of four positions around a fixation point: above, below, left, or right. In each of the experiments, *S*s were run under conditions in which they could not see the positions of their upper limbs. Compatibility effects were found in both experiments, showing that such effects can occur when the positions of the responding members are specified by kinesthetic cues alone. The results also suggest that the coding of stimulus and response position was in terms of left and right rather than specific positions along the horizontal.

Previous experiments have shown that in a situation in which the response keys are placed to the left and right of *S* and the cue stimuli are presented to his left and right as well, the reaction time is less when the cue stimulus and the responding member are on the same side than when they are on opposite sides. This compatibility effect is found even though the position of the stimulus is not a relevant cue for the response (Craft & Simon, 1970; Simon & Rudell, 1967; Simon & Small, 1969). In addition, these relationships among cue stimulus, responding hand, and reaction time, are found whether or not *S*'s hands are crossed (Simon, Hinrichs, & Craft, 1970; Wallace, 1971). This indicates that the relation between the position of the cue stimulus and the particular part of the body used to make the response is not important for this effect; rather, it is the position of the cue and the position of the responding member that determines the difference in reaction time. This, in turn, suggests that the cue and the responding member should both be considered as inputs which are then analyzed. When the positions of the cue and the hand are similar, this analysis results in the facilitation of the motor output related to that hand (for further discussion of these ideas see Wallace, 1971).

The present experiments are similar to the previous experiment by this author (Wallace, 1971), except that *S*s could not see their limbs at any time during the experiment. Therefore, information about the position of the responding member must be derived from somatosensory or motor systems. If the same effects are found here as in the previous experiment, it can be concluded that position information from these systems can be correlated with information from the visual system to produce these compatibility effects.

In addition, these experiments provide evidence about another characteristic of the hypothesized spatial code which was left ambiguous in earlier compatibility experiments. In the experiment by Wallace (1971), for example, the response keys were directly beneath the left and right positions of the cue stimuli. Therefore, it is not clear to what degree the spatial coding was in terms of specific positions along the horizontal, or in terms of intervals along this dimension. In the Wallace experiment, the simplest form of interval coding would have been in terms of two intervals, i.e., all those positions to the left or to the right of some reference point, e.g., *S*'s midline. In this case, as long as the positions of the cue stimulus and the responding member were both on the left, or the right, the response would be facilitated. In the present experiments, the cue stimuli are identical to those used previously (Wallace, 1971), while the response keys are farther apart. Thus the positions along

---

[1] Requests for reprints should be sent to Richard Wallace, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

TABLE 1

Mean Reaction Times in Milliseconds

| Cond. | Response | Position of figure | | | |
|---|---|---|---|---|---|
| | | Left | Right | Top | Bottom |
| Kinesthetic cond. | | | | | |
| Hands uncrossed | Left key | 422 | 462 | 452 | 456 |
| | Right key | 477 | 414 | 439 | 441 |
| Hands crossed | Left key | 448 | 499 | 450 | 480 |
| | Right key | 519 | 457 | 494 | 471 |
| Nonkinesthetic cond. | | | | | |
| Hands uncrossed | Left key | 396 | 455 | 430 | 414 |
| | Right key | 421 | 382 | 411 | 401 |
| Hands crossed | Left key | 439 | 489 | 484 | 464 |
| | Right key | 496 | 455 | 471 | 469 |

the horizontal are no longer the same, although they are still on the left and right. Therefore, a comparison of the size of the compatibility effects in the two experiments may indicate the degree to which these different types of coding are involved.

## Experiment I

### *Method*

*Subjects.*—Twenty-four male and female students of the University of Oregon served as *S*s. All were right-handed and had at least 20–20 vision. Each *S* was paid $1.50 for the session. Upon being questioned at the end of the session, all but 1 of the *S*s indicated that they were not aware of the purpose of the experiment. The 1 *S* was included anyway; his data did not differ obviously from that of the others.

*Materials and apparatus.*—These were similar to those of a previous experiment (Wallace, 1971: Exp. I) and involved only two changes. The first was that the response keys (3 cm. wide) were placed about 27 cm. apart, or about 15 cm. on either side of *S*'s midline. The second change is discussed under the Procedure section.

*Procedure.*—The procedure was similar to that of the previous experiment. On each trial, *S* fixated on a dot in the center of the field of a tachistoscope. About 1 sec. after the warning signal, given by *E*, a figure, either an outline circle or square, was presented for 80 msec., either to the left or the right or above or below the fixation point. If the figure was a circle, *S* had to press the left of two keys placed before him; if it was a square, he had to press the right key.

Each *S* was run for 1 session of 128 trials. The *S*s were run in a 2 × 2 design with six *S*s in a group. The first dimension of the design had to do with whether *S* was run with his hands uncrossed, so that his left and right hands were on the left and right keys, respectively, or with his hands crossed so that his right hand was on the left key, and vice versa. In both of these conditions, *S*s pressed the response keys down with their middle fingers. For the hands-crossed conditions, three *S*s were run with their right hand over their left and three were run in the opposite fashion. The second dimension had to do with whether *S* could see his hands on the response keys between trial blocks. The following procedure was used to prevent *S* from seeing his hands. During the experiment, *S* wore a halter made from a piece of cardboard about 3 ft. wide, with aluminum straps, which were placed over his shoulders. A piece of cloth also 3 ft. wide was cut at one end and stapled onto the cardboard so that it fit around *S*'s neck; the other end was attached to the table on which the tachistoscope rested. Thus, when *S* had the halter on and was seated before the tachistoscope, he could not see any part of his upper limbs or the response keys. In order to get *S* into the halter without him seeing the response keys, he was brought into the experimental room blindfolded; in addition, the room was darkened and the halter was placed over the response keys. While wearing the halter, *S* was able to lean forward to look into the tachistoscope and to lean back in the chair during the break between blocks of trials; the length of the cloth between its attachments was, of course, made short enough that it did not rest on his hands when he was leaning forward. The two conditions run in this way will be called the kinesthetic conditions. The two other conditions, in which *S*s were run without the halter and could therefore see their hands, will be referred to as the nonkinesthetic conditions.

*Data analysis.*—A repeated-measures analysis of variance was used to analyze the results. The kinesthetic and nonkinesthetic conditions were analyzed separately. In addition, the left and right positions of the cue stimuli were analyzed separately from the top and bottom positions. Analyses were made for both reaction times and errors, so that there was a total of eight ANOVAs. In each analysis there were four factors: *S*s, position (of the stimuli: either left and right or top and bottom), key (left or right), and condition (hands uncrossed or crossed). Error reaction times were not included in the analysis, nor were reaction times over 1.5 sec. In the kinesthetic hands-crossed condition there was one such reaction time, and in the nonkinesthetic hands-crossed condition there were four.

### *Results*

The results of this experiment are given in Table 1. The compatibility effects were found in all four conditions and were about the same size in each.

The analyses of variance will be discussed beginning with the reaction time ANOVAs and the kinesthetic conditions. In each ANOVA, the term which is related to the compatibility effect is the Position × Key interaction. For the left–right ANOVA, the only significant effect was the Position × Key interaction, $F\ (1,\ 5) = 38.8$, $p < .005$. For the corresponding top–bottom ANOVA, only the Position × Key × Condition interaction was significant, $F\ (1,\ 5) = 7.3$, $p < .05$. This is because a compatibility effect was also found for the top and bottom positions in the hands-crossed condition (see Table 1 and Discussion section).

In the nonkinesthetic conditions, the following effects were significant in the left–right ANOVA: Position × Key, $F\ (1,\ 5) = 240.7$, $p < .001$, condition, $F\ (1,\ 5) = 6.6$, $p < .05$, and Key × Condition, $F\ (1,\ 5) = 7.2$, $p < .05$. The significance of the Position × Key and condition effects are related to the compatibility effect and the slower responses in the hands-crossed condition, respectively. The Key × Condition interaction was significant because the right *hand* was faster whether or not the hands were crossed. For the top–bottom ANOVA, the effect of condition was significant, $F\ (1,\ 5) = 12.0$, $p < .025$.

In the error ANOVA, for the kinesthetic conditions, the Position × Key interaction was significant in both the left–right and top–bottom analyses, $F\ (1,\ 5) = 34.1$, $p < .005$, and $F\ (1,\ 5) = 12.0$, $p < .025$, respectively. For the nonkinesthetic conditions, the Position × Key interaction was significant in the left–right ANOVA, $F\ (1,\ 5) = 35.1$, $p < .005$, and there were no significant effects for the top–bottom analysis. In each of the four conditions, the error rate was 5% or less.

In all four conditions, as in the earlier experiment (Wallace, 1971), for each response the mean reaction times to the top and bottom positions fell in between those to the left and right. In the previous experiment, it was suggested that such a finding indicates that there is both facilitation and interference in the compatible and incompatible conditions, respectively. The present results indicate that this is also the case when *S* cannot see his hands.

## EXPERIMENT II

In the previous experiment, in the kinesthetic conditions, *S*s located the response keys by moving their hands until they felt the keys. It is conceivable that information about the position of the responding member was provided by such movements, through reafferent input, or corollary discharge (efference copy) from motor command units at some level in the nervous system (cf. vonHolst, 1954). That is, since there is a pattern of input from muscle and joint receptors which is correlated with the movement to the response keys, and since a pattern of activity in a set of motor command units is also correlated with such movement, the information contained in either or both could be used in determining what movement had been made and what end position had been reached. The first two conditions of Exp. II were run to determine whether these active movements did have an influence on the observed compatibility effects. In the third condition, there was one further alteration in that the instructions as to how to respond were stated in terms of the hand that was to press the key rather than the key that was to be pressed down. Although such compatibility effects have been demonstrated using "hand" instructions (Simon et al., 1970), it was felt that it was important to determine the possible effect of "key" instructions in the present situation, where the cues for the position of the responding member were minimal.

### *Method*

*Subjects.*—Twelve male and female students of the University of Oregon served as *S*s. All were right-handed and had 20–20 vision. Each was paid $1.50. Questioning revealed 1 *S* who had guessed the purpose of the experiment.

*Materials and procedure.*—These were similar to those of the kinesthetic conditions of Exp. I, with the following differences. After the instructions were given, *S*'s hands were moved one at a time by *E* from a midline position to the appropriate response key; while this occurred, *S* was instructed to keep his arm relaxed. Once in position, the arm was strapped down, so that during the experiment his arms could not move more than a few centimeters in any direction. The *S*s responded with the forefinger of each hand. Four *S*s were run in each

TABLE 2

MEAN REACTION TIMES IN MILLISECONDS

| Cond. | Response | Position of figure | | | |
|---|---|---|---|---|---|
| | | Left | Right | Top | Bottom |
| Hands uncrossed: "key" instructions | Left key | 378 | 449 | 432 | 415 |
| | Right key | 446 | 384 | 404 | 401 |
| Hands crossed: "key" instructions | Left key | 451 | 500 | 487 | 492 |
| | Right key | 526 | 467 | 485 | 499 |
| Hands crossed: "hand" instructions | Left key | 432 | 477 | 455 | 469 |
| | Right key | 490 | 475 | 480 | 485 |

of three conditions: hands-uncrossed key instructions, hands-crossed key instructions, hands-crossed hand instructions. In the key instruction conditions, as in previous experiments, *S*s were told to press the left key if the figure was a circle and the right key if it was a square. In the hand instruction condition, *S*s were told to respond with their right hand if the figure was a circle and with the left hand if it was a square. In the hands-crossed conditions, all *S*s were run with their left hand over their right.

*Data analysis.*—The key conditions were analyzed together, the hand condition separately, in a fashion similar to Exp. I. Three reaction times, from one *S*, in the key hands-crossed condition were over 1.5 sec. and were not included in the analysis.

### *Results*

The results of this experiment are given in Table 2. The compatibility effects for the left and right positions were found in all three conditions. In the analyses of variance for the key conditions, the Position × Key interaction was significant for the left–right analysis, $F\ (1,\ 3) = 77.3$, $p < .005$, and the Key × Condition interaction was significant for the top–bottom analysis, $F\ (1,\ 3) = 25.6$, $P < .025$, for the reasons discussed before. For the hand condition, the Position × Key interaction was significant in the left–right ANOVA, $F\ (1, 3) = 43.8$, $p < .01$. In addition, the key effect was significant, $F\ (1, 3) = 170.3$, $p < .001$, because the left key (right hand) was faster than the right key. There were no significant effects in the top–bottom ANOVA.

The only significant effect in any of the error ANOVAs was the Position × Key interaction in the left–right analysis for the key conditions, $F\ (1,\ 3) = 49.0$, $p < .01$. In the hand condition, the errors showed the expected pattern, but the Position × Key term was not quite significant, $F\ (1, 3) = 9.0$, $.05 < p < .10$. In each condition, the error rate was 4% or less.

Again, the mean reaction times for each response to the top and bottom positions were between those for the compatible and incompatible S–R pairings.

## DISCUSSION

The experiments show that the compatibility effects are found in this situation even when *S* cannot see his hands. Experiment II seems to rule out any form of reafference or corollary discharge during active movement as necessary for the effects. Therefore, information about the position of the responding member must be derived from input to somatosensory receptors. Tactile inputs may have served as cues in the hands-crossed conditions, in which one arm was placed over the other. However, it is questionable whether such inputs contain enough information to specify the position of these parts of the body accurately, or whether this simultaneous stimulation of both limbs is sufficiently unambiguous to be used as a cue for position. Finally, such information is not available in the hands-uncrossed conditions, but the effects found in these conditions were similar to those found with the hands crossed. Therefore, it would seem that tactile inputs provided little or no essential information about the positions of the responding members in these experiments. If this is true, then the necessary information must have been transmitted over kinesthetic pathways, presumably from receptors associated with the joints (see Brindley & Merton, 1960; Provins, 1958; Rose & Mountcastle, 1959).

Thus the experiments show that the similarity in position of the cue stimulus and responding member, which results in facilitation or interference, can involve a correlation of information from visual and kinesthetic modalities. At some stage, inputs to both modalities evidently activate symbols in a common code, which is related in some way to positions in space, and the operations which underlie the compatibility effects must then involve the results of this coding process. This conclusion is also suggested by the results of Szafran (Welford, 1958), who used a roughly similar arrangement to prevent *S* from seeing his hands. These data may be taken as complementary to experiments in which information about the position of the cue stimulus is transmitted via auditory pathways and then correlated with information from other modalities (e.g., Simon et al., 1970; Simon & Small, 1969).

Since the nonkinesthetic conditions in Exp. I were similar to those of Day 1 of the first experiment in Wallace (1971) in all respects, except the positions of the response keys, a comparison of the sizes of the compatibility effects in these two experiments can provide evidence on the question of position versus interval coding. In order to compare the sizes of the compatibility effects in the same way under kinesthetic conditions, a further experiment was run in which the two keys directly below the cue stimulus positions were used, as in the Wallace experiment; otherwise, it was identical to the kinesthetic conditions of Exp. I in the present study. Analyses of variance were performed separately for the kinesthetic and nonkinesthetic and hands-uncrossed and hands-crossed conditions, so that four were done in all; the effect related to the difference in the placement of the response keys will be called placement. For three of the ANOVAs, the relevant effect, Position × Key × Placement, was not significant, $F\ (1,\ 5) \leq 1.0$, $p > .20$. For the fourth it was, $F\ (1,\ 5) = 6.8$, $p < .05$. This was because, in this condition (kinesthetic, hands crossed), the compatibility effects were somewhat smaller for the keys directly beneath the cue stimuli, perhaps because the difference in position was less salient under these conditions. The analyses suggest, then, that in the present situation the coding of the positions of the cue stimuli and responding members was soley in terms of intervals. It is doubtful that this is a code of positions with poor resolution, since the outer and inner keys were about 11 cm. apart, which is a clearly discriminable difference in position. The data are consistent with the idea that in a situation with two responding members in left and right positions, the coding of position along the horizontal which results in compatibility effects is in terms of left and right.

Since the compatibility effects are related to whether the cue stimulus or responding member is on the left or right, it is conceivable that the code in question is a name code, consisting of the names "left" and "right." Thus, if both the cue and responding hand are on the left side, they may both be classified as "left"; and the similarity in classification could result in the facilitation of the response. However, an experiment by Simon and Rudell (1967) suggests that this is not the case. In this experiment, the cue stimuli were the words "left" and "right," presented in either the left ear or the right ear. The *S*s had to press the left of two keys when the word was "left," and the right key for "right." Compatibility effects were found between the ear stimulated and the key pressed; e.g., the left key was faster when the word "left" was presented in the left ear. However, these cue stimuli should be classified as "left" or "right" regardless of their position, because this is the way they are naturally coded. Under the hypothesis, the responding members are also coded in the same way in this situation, so the classification should always be similar. Hence, with these cue stimuli there should be no compatibility effects. I am assuming, of course, that compatibility effects in the present situation involve the same coding processes as those studied by Simon and Rudell, but there seems no reason to doubt this. Therefore, the code involved here is really a spatial code, but it is "degraded" so that there are only two values (along the horizontal).

In addition to the left–right compatibility effects, some of the data from the hands-crossed conditions gives evidence for a top–bottom compatibility effect. That is, there was some tendency to code both the cue stimuli and the limbs involved in responding as on top or on the bottom, so that the limb on top was faster when the cue was above than when it was below the fixation point, while the limb on the bottom was faster when the cue was below. However, these effects were not consistently present, nor was there a consistent relation between similarity in position and faster reaction time. This inconsistency is, in fact, the reason for the significant Position × Key × Condition interaction found for the top–bottom analysis in Exp. I, kinesthetic condition; there, *S*s who were run with their left hand over their right performed in a manner opposite to that expected (in the nonkinesthetic condition there were no reversals, but a difference was not always present). In general, this effect seemed to be more consistent when the right hand was over the left.[2]

It should be noted that in this situation, above and below must refer to a different reference position in the case of the cue stimuli than in the case of the responding members, since both cue positions were above the positions of the two hands. A similar finding has been made by Frost and Snyder (1970) for left–right compatibility effects. They found that if the response keys lie along the midsagital axis and the cue stimuli along the horizontal, the key nearest *S* is faster when the stimuli are on the left, and similarly for the far key and stimuli on the right (*S* pressed the keys with two fingers of his right hand). Thus, it appears that even when the cue stimulus and response positions form different arrays, positions in similar parts of their respective arrays can be compared, to produce compatibility effects.

## REFERENCES

Brindley, G. S., & Merton, P. A. The absence of position sense in the human eye. *Journal of Physiology*, 1960, **153**, 127–130.

Craft, J. L., & Simon, J. R. Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology*, 1970, **83**, 415–420.

Frost, R. R., & Snyder, C. R. R. Stimulus localization and attention in a reaction time paradigm. Paper presented at the Meeting of the Western Psychological Association, Los Angeles, April 1970.

Provins, K. The effect of peripheral nerve block on the appreciation and execution of finger movements. *Journal of Physiology*, 1958, **143**, 55–67.

Rose, J. E., & Mountcastle, V. B. Touch and kinesthesis. In J. Field (Ed.), *Handbook of physiology, section 1, neurophysiology, vol. 1*. Washington, D. C.: American Physiological Society, 1959.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. Auditory S–R compatibility: Reaction time as a function of ear-hand correspondence and ear-response-location correspondence. *Journal of Experimental Psychology*, 1970, **86**, 97–102.

Simon, J. R., & Rudell, A. P. Auditory S–R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 1967, **51**, 300–304.

Simon, J. R., & Small, A. M., Jr. Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 1969, **53**, 433–435.

vonHolst, E. Relations between the central nervous system and the peripheral organs. *Animal Behavior*, 1954, **2**, 89–94.

Wallace, R. J. S–R compatibility and the idea of a response code. *Journal of Experimental Psychology*, 1971, **88**, 354–360.

Welford, A. T. *Ageing and human skill*, Oxford University Press, 1958.

(Received August 24, 1971)

---

[2] In Exp. I of Wallace (1971), top–bottom compatibility effects were also present; the size of the effects ranged from 25 to 43 msec. The same effects, although smaller, were found under kinesthetic conditions with the inner keys. Thus the effect also seemed to be more consistent when response keys 7 cm. apart were used, perhaps, because in this case one hand was on top of the other hand (rather than one forearm on top of the other forearm). In order to determine whether there is a significant effect on reaction time in the hands-crossed condition to cues in the top and bottom positions, two groups of six *S*s were run: one with the right hand over the left, the other with the left hand over the right. The *S*s were run for one session with the two inner response keys. Top–bottom compatibility effects were found in both groups; the average effect in both cases was about 25 msec. For the right-over-left group the effect was significant (for the Position × Key interaction, $F\ (1, 5) = 40.5$, $p < .005$); but for the left-over-right group it was not, $F\ (1, 5) = 3.5$, $p > .10$

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 169–175

# LATENCY OF LOCATING LIGHTS AND SOUNDS[1]

W. E. SIMPSON[2]

*State University of New York, College at Oswego*

As a complement to past research on the accuracy of visual and auditory spatial localization, three experiments investigated localization latency. Stimulus azimuth and auditory intensity were varied as was response mode—key pressing or verbalization ("right," "left"). Reaction time (RT) for localization was compared against simple RT control data. Analyses of mean RT and variablitiy of RT supported the hypothesis that auditory localization processes are more complex than visual localization processes. Giving *S* foreknowledge of the modality of a stimulus affected simple and localization RTs equally, indicating that stimulus uncertainty did not affect localization per se.

Objects can be localized in space more accurately using the visual perceptual system than using audition. For example, the error in pointing one's unseen finger to a spot of light is less variable than when pointing to a sound. Warren (1970) had adult *S*s point to lights and sounds coming from a variety of azimuths. The standard deviation of the error of response was about 1.5° for visual targets versus about 5.0° for auditory targets. Inducing *S*s to look in the direction of an invisible auditory target before pointing resulted in improved performance, although this performance was still inferior to localization of visual targets. Pick, Pick, and Klein (1967) summarize control data from several similar studies which further support the generalization that vision is a more precise system for localization. In a quite different setting, Rekosh and Freedman (1967) studied the effects of wearing a wedge prism upon the ability to align visual or auditory stimuli with the midline. Comparing pretest versus posttest and trial-by-trial responses, the alignments made by control *S*s, who did not wear prisms during "adaptation," were more variable for auditory than for visual targets.

One interpretation of these studies is that localization via the visual system involves less complex processing than auditory localization. The term complex is here used very generally. But Warren (1970) has shown one way in which the concept might be usefully conceived. If auditory localization is dependent upon the concomitant use of visual information, i.e., Warren's "visual map" or reference system, then auditory localization is necessarily more complex than visual localization.

Our understanding of the processes involved in visual and auditory localization should be improved by considering dimensions other than precision of pointing. For example, a comparison of the times required to locate lights and sound should give some indication of the relative complexities of the respective processes. Greater complexity of processing could be expected to be reflected in either or both of two ways, greater response latency or greater variability of response latencies. Perhaps the simplest example of localization processing would require *S* to indicate whether a stimulus had occurred on his left or his right. This was the task examined in the present study. Reaction time (RT) in this localization task was then compared with simple RT, for which *S* was required to make a predetermined response to the onset of the same stimuli used in the localization task.

One of the standard facts in the repertory of psychology is that simple RT is about 40 msec. less to an auditory than to a visual stimulus (Woodworth & Schlosberg, 1954). Recent experiments by Kohfeld (1971) have shown this to be an overgeneralization. Kohfeld suggested that when visual and auditory stimulus intensities are adjusted to levels that correspond to comparable positions on their respective sensory continuua, visual and auditory stimuli may be responded to with equal speed. Thus, when a white light and a 1,000-Hz. tone were both set at 60 db. above their respective absolute thresholds, Kohfeld found that simple RT was equal for the two stimuli. But a modality difference did occur when visual stimuli were below the photopic threshold. That is, with stimuli set at 30 db. RT to the auditory stimulus was faster than to the visual stimulus. The factor critical for producing a modality difference in RT is therefore whether the visual stimulus activates cones or rods. An implication of Kohfeld's study is that the traditional modality difference in RT must reflect the use of experimental conditions in which either the stimuli were at scotopic (rod) levels or the visual stimulus was at a lower sensation level than the auditory stimulus. In the present study, visual stimuli were located far enough in the periphery to stimulate primarily rods (see Geldard, 1953, p. 25) and therefore a modality difference in simple RT was predicted.

---

[1] This research was supported by United States Public Health Service Grant MH 18283-01 and by the State University of New York Research Foundation. Martha Hare collaborated on Exp. III through support of the National Science Foundation Undergraduate Research Participation program.

[2] Requests for reprints should be sent to W. E. Simpson, Department of Psychology, State University of New York, College at Oswego, Oswego, New York 13126.

There has long been controversy over whether it is valid to assume that the various component processes contributing to RT are combined additively (LaBerge, 1971). In the present study, only an elementary and defensible version of additivity need be assumed: that a stimulus must activate a receptor before it can be localized. The increment in localization RT relative to simple RT should therefore reflect the localization process. Intermodal comparison of localization RTs versus simple RTs should then indicate the relative complexities of auditory and visual localization. That is, if the modality difference observed for simple RT is eliminated or reversed in the localization task, auditory localization is indicated as the more complex. If the modality difference is increased, the visual localization process is the more complex. In line with Warren's (1970) results, above, the former relationship was predicted here. That is, when progressing from detection (simple RT) to localization, a greater increment in RT was predicted for auditory than for visual stimuli. The hypothesized difference in complexity of auditory and visual localization also leads to a prediction about the variability of RT responses. If localization processing is more complex for auditory than for visual stimuli, then the variability of localization responses should be greater for auditory than for visual stimuli. No such difference was predicted for simple RTs.

## EXPERIMENT I

The major procedural problem faced in comparing the times required to localize visual and auditory stimuli is the establishment of comparable stimuli in the two modalities. A useful statement of intermodal differences in localization should hold over a range of stimulus characteristics, e.g., specific locations and specific stimulus intensities. Experiment I sought to establish simple and localization RTs at two loci, 30° and 60° off the midline. The 30° position was selected because it seemed to offer an easy right–left discrimination for both audition and vision. The 60° locus was chosen to give a possible advantage to localizing the auditory stimulus. That is, right–left discriminations are made with less error as the stimuli are moved more lateral while at the same time visual stimuli become less detectable (and less bright) in the periphery. Thus, if comparable intermodal effects are found for stimuli at both loci, the interpretation of intermodal differences in localization processing time will be shown to have generality. Comparison of stimuli located more medially than 30° would be of little value, since here an obvious advantage accrues to the visual stimuli.

*Method*.—The *S*s were 24 student volunteers (12 males) from introductory psychology classes. Twelve *S*s (6 males) were randomly assigned to each of two groups, one performing a simple RT task and the other a localization RT task. The *S* sat facing the interior surface of a black-cloth hemisphere with radius 90 cm. A small fixation point was visible at eye level directly in front of *S*. The room was dimly illuminated but there were sufficient cues to preclude any loss of spatial orientation. Background noise registered 55 db. (C scale, General Radio 1551-C meter). Four small 110-v. neon lamps (GE F3A) served as the visual reaction stimuli, located at 30° and 60° left and right of the fixation point. Auditory stimuli were 800-Hz. bursts presented through four 2-in. speakers placed just beneath the lights. All stimulus sources were behind the cloth screen and were presented for a duration of 50 msec. Using the 30° positions, auditory stimulus loudness was matched to the visual stimulus brightness by 5 independent *S*s and the mean matched sound level, 60 db., was maintained throughout the experiment. Experimental *S*s were instructed to focus on the fixation point when a verbal "ready" signal was given before each trial. A variable foreperiod of 1–2 sec. was used. Eight practice trials were given under the appropriate condition (simple or localization RT).

In the simple RT group, *S*s were asked to respond as quickly as possible by verbalizing the arbitrary response "left" ("right"). Each response was used by half of the *S*s. An *S* was allowed to discard trials on which he judged himself unready, although this was rarely done ($< 1\%$ of the trials). RTs were collected for 15 randomized replications of the eight possibilities (light 30° left, sound 30° left, etc.), a total of 120 trials for each *S*.

Stimuli and procedures were identical for *S*s in the localization RT group with the following exceptions. An *S* was required to respond "left" or "right" depending on the stimulus locus, and was asked to respond fast enough to hold an error rate of about 5%. Erred trials were replaced later in the sequence. Thus, the two experimental conditions were identical in terms of the stimuli and responses involved and differed only in terms of the nature of the intervening decision required of *S*.

Voice responses were transmitted by a throat microphone to a Grayson-Stadler E7300M voice-operated relay which stopped a Hunter 1520 counter which had been started at the onset of the reaction stimulus. RTs were recorded to the nearest millisecond.

*Results*.—In the localization condition, the mean error rate was 3% and did not differ for auditory and visual stimuli. At the conclusion of the experiment, *S*s were asked whether it had been easier to respond to one stimulus mode. Only one *S*, in the localization group, indicated that the visual stimulus had been easier; most *S*s in both groups said that the auditory stimulus had been easier. These indexes suggest that the auditory stimulus was certainly not disadvantaged in present setting.

TABLE 1

MEAN REACTION TIME AND STANDARD DEVIATION IN MILLISECONDS FOR EXPERIMENT I

| Mode | Simple RT | Localization RT | Both tasks |
|---|---|---|---|
| Auditory | | | |
| 30° Azimuth | 415(61) | 636(116) | 526(89) |
| 60° Azimuth | 395(61) | 610(103) | 503(82) |
| Both Azimuths | 405(61) | 623(110) | 514(86) |
| Visual | | | |
| 30° Azimuth | 431(51) | 615 (96) | 523(74) |
| 60° Azimuth | 458(64) | 617 (87) | 538(76) |
| Both Azimuths | 445(58) | 616 (92) | 531(75) |
| Overall | 425(60) | 620(101) | 523(81) |

Note.—Standard deviations are given in parentheses.

A mean RT was calculated for each azimuth-modality combination, collapsing right and left sides, for each *S*. These means were subjected to a repeated-measures analysis of variance. Group means are presented in Table 1. The standard error of a mean, estimated from the error mean square of the analysis of variance, was 11 msec. Significant sources of variance were task (simple RT less than localization RT), $F(1, 20) = 27.5, p < .01$, stimulus mode (faster responding to the auditory stimulus), $F(1, 60) = 9.4, p < .01$, and the Task × Mode interaction, $F(1, 60) = 18.0, p < .01$. A Scheffé test (Edwards, 1960) of the interaction indicated that simple RTs were shorter to the auditory than to the visual stimuli (405 vs. 445 msec.) but that there was no difference between stimulus modes in the localization task (623 vs. 616 msec.). Azimuth was not a significant source in the analysis of variance. Visual inspection (Table 1) indicated that RTs to auditory stimuli tended to be less at 60° than at 30° but that visual RTs tended to be less at 30° than at 60°. This pattern is consistent with the logic used in establishing the loci for the experiment. Yet, at the level of precision of this study, the azimuth variable is not an appreciable contributor to variance, although task and mode are. This confirms the generality of the fact that the relative superiority of the sound as a stimulus for *detection* is lost under the added requirement of *localization*.

To test the hypothesis of greater response variability when localizing sounds, the standard deviation was calculated for each *S*'s responses to auditory and to visual stimuli at the 30° and at the 60° positions. These standard deviations were subjected to analysis of variance, and group mean standard deviations are included in Table 1. Standard deviations were greater for localization than for simple RTs, $F$ (1, 20) $= 5.0$, $p < .05$, they were also greater for auditory than for visual stimuli, $F$ (1, 60) $= 9.7$, $p < .01$. In confirmation of the above hypothesis, the interaction of these variables was also significant, $F$ (1, 60) $= 4.7$, $p < .05$. It is evident that response variability of simple RTs is the same for auditory and visual stimuli (even though the mean RTs are different) but that localization RTs are more variable for auditory than for visual stimuli (even though the mean RTs are not different). The Task × Azimuth interaction was also significant, $F$ (1, 60) $= 6.4$, $p < .05$, and indicated that detection (simple RT) was less variable for the more medial stimulus locus whereas localization was less variable for the more lateral locus.

## EXPERIMENT II

In the first experiment the subjective intensities of the visual and auditory stimuli were set by a cross-modal matching procedure. Intuitively this seemd to provide a "fair" comparison of processing in the visual and auditory systems. But how sensitive is the inferred difference in localization processing time to variation in stimulus intensity? To the extent that observed RTs truly reflect slower auditory localization processing, this inferiority should be maintained under increased auditory stimulus intensity. The Kohfeld (1971) study demonstrated that observed modality differences in simple RT were peripheral in origin. Stimulus intensity was interpreted as influencing simple RT through the rate at which a stimulus reached detection threshold. Localization processing is presumably more centrally mediated and less sensitive to stimulus intensity. Thus variation in auditory stimulus intensity should produce equal changes in both simple and localization RTs to the auditory stimulus, relative to the visual stimulus.

Experiment II was therefore designed to test the generality of the pattern of results from Exp. I under conditions where auditory stimulus level was cross-modally matched to the visual stimulus, or higher, or lower. It was predicted that an increase in auditory stimulus intensity would speed up both simple and localization RT by the same amount.

*Method.*—The *S*s were 24 student volunteers from the same pool and assigned to groups as in Exp. I. Apparatus and stimuli were identical to Exp. I, with the exception that the auditory stimulus level was varied from trial to trial in an unpredictable fashion. Three intensity levels were used equally often, 55 db., 70 db., and an intermediate level. The latter was established for each *S* at the beginning of the session as the mean cross modal match from three each ascending and descending series of trials using the method of limits. The mean intermediate level was 61 db. and did not differ for the simple and localization RT groups. Stimuli were located at 30° right and left of the fixation point. Each *S* was presented with eight different replications of 12 randomized trials. Each replication consisted of 6 trials to an auditory stimulus (1 each of the 3 intensities and 2 loci) and 6 trials to visual stimuli (3 at each locus).

*Results.*—Means and standard deviations of RTs are presented in Table 2. Standard error of a mean was 10 msec. In the cases of both simple and localization RTs, each increment in auditory stimulus intensity decreased RT by about 10 msec. Although these numerical changes were in the expected direction, a Scheffé test indicated

TABLE 2

MEAN REACTION TIME AND STANDARD DEVIATION IN MILLISECONDS FOR EXPERIMENT II

| Mode | Simple RT | Localization RT | Both tasks |
|---|---|---|---|
| Auditory | | | |
| 55 db. | 422 (50) | 579 (90) | 501 (70) |
| Matched | 414 (52) | 570 (84) | 492 (68) |
| 70 db. | 399 (58) | 563 (85) | 481 (72) |
| Overall | 412 (53) | 571 (86) | 492 (70) |
| Visual | 455 (41) | 576 (65) | 516 (53) |
| Both | 434 (47) | 574 (76) | 504 (62) |

Note.—Standard deviations are given in parentheses.

that they were not statistically significant. The Task × Modality interaction was significant, $F$ (1, 100) = 21.0, $p < .01$. A Scheffé test indicated that simple RT responses were faster to the auditory than to the visual stimulus (412 vs. 455 msec.). The test showed no difference in localization RT as a function of modality (571 and 576 msec.). Again, these relationships held irrespective of auditory stimulus level.

As in Exp. I, standard deviations were also analyzed. Standard deviations were greater for localization RT than for simple RT, $F$ (1, 20) = 23.8, $p < .01$, and greater for auditory stimuli than for visual stimuli, $F$ (1, 20) = 21.5, $p < .01$. The task and modality variables did not interact in affecting response variability, $F$ (1, 20) = 1.56.

Error rate in the localization task averaged 1% for both auditory and visual stimuli. As in Exp. I, questioning *S*s after the experiment suggested that responding to auditory stimuli was judged no more difficult than responding to visual stimuli, regardless of the task.

Taken together, Exp. I and II demonstrate that the inferred slower localization processing of auditory information is real, holding for stimuli at both 30° and 60° azimuth and through a 15-db. range of auditory stimulus intensity.

Two other questions can be raised about the conclusion that auditory localization is a slower process than is visual localization. The first is a question about the validity of attributing the observed differences in "localization RTs" to localization per se. Experiments I and II used a verbal response. In most experiments on reaction time the response mode is manual (e.g., key-pressing). With reference to the present localization task, it might be argued that a right–left manual response would provide a more direct and valid measure of localization. There are two reasons for this: (*a*) a series of experiments by Simon (1969) has demonstrated that *S*s in RT experiments are indeed biased in favor of making manual responses in the direction of a stimulus, and (*b*) one notion about the functional significance of localization processes is that they serve to orient *S*'s receptor surfaces for more fine-grain analysis of stimulus information—and manual pointing implies orientating. If *S* were required to indicate the apparent locus of a peripheral reaction stimulus by pressing a key ipsilateral to the stimulus, this would therefore seem to provide a response quite different from the verbal mode, and, quite fair or valid structurally. It therefore becomes of interest to try to replicate the findings of Exp. I and II using a manual (key press) response.

The second question concerns the flexibility of localization processing. If *S* were able to predict that a stimulus would be in the auditory mode would he be able to alter the amount of time required to process that stimulus? Experiment III was designed to compare simple and localization RTs under conditions where *S* was given or not given foreknowledge of the stimulus modality. It was predicted that foreknowledge of stimulus mode would speed up simple RT by some small fixed amount. Such a prediction follows if mode predictability is considered as an example of reduction of stimulus uncertainty, which has been shown to reduce RT (Smith, 1968). Prediction about the influence of foreknowledge on localization is more difficult. Foreknowledge might have no effect other than that observed in simple RTs. Or there might be a speeding up of auditory localization, but not visual localization. This could be interpreted in several ways, e.g., as indicating that auditory localization is not subservient to a visual reference system but that under mode uncertainty *S* is merely biased in favor of processing visual stimuli. It might also be interpreted as indicating that visual referencing can (but typically does not) occur simultaneously with other aspects of auditory localization. At its current level of development, the notion that vision serves as a referent for auditory localization does not suggest a clear choice between these outcomes or interpretations.

## EXPERIMENT III

*Method.*—Ten *S*s (5 males) were randomly assigned to each of four groups. The *S*s ranged from

TABLE 3

MEAN REACTION TIME AND STANDARD DEVIATION IN MILLISECONDS FOR EXPERIMENT III

| Mode | Simple RT | Localization RT | Both tasks |
|---|---|---|---|
| Auditory | | | |
| Predictable | 250(29) | 310(47) | 280(38) |
| Unpredictable | 309(50) | 347(50) | 328(50) |
| Overall | 280(40) | 329(48) | 304(44) |
| Visual | | | |
| Predictable | 316(31) | 339(38) | 328(35) |
| Unpredictable | 355(45) | 373(41) | 364(43) |
| Overall | 336(38) | 356(40) | 346(39) |
| Both | | | |
| Predictable | 283(30) | 325(43) | 304(37) |
| Unpredictable | 332(47) | 360(46) | 346(47) |
| Overall | 308(39) | 343(44) | 325(42) |

Note.—Standard deviations are given in parentheses.

18 to 30 yr. of age and had normal hearing and vision (corrected). Two of the groups performed a simple RT task. For one group, both stimulus mode and locus (30° left, right) were determined randomly from trial to trial, with the restriction that there be 16 trials each: visual right, left; auditory right, left. Trials for the second group were presented in blocks of 16 trials of a given stimulus mode. These modality blocks were alternated (VAVA or AVAV) and stimulus locus was ordered randomly within each block. Thus, these two groups differed only in the extent to which stimulus mode was predictable on a given trial (50% vs. 100%). Prior to these experimental trials, all *S*s performed 8 unscored practice trials and then 20 mode-predictable warm-up trials which provided scores for testing the equivalence of groups on overall RT. The third and fourth groups performed a localization RT task, one with mode unpredictable and the other with mode predictable. The only difference between the simple RT and the localization RT groups was in the response requirement. Localization RT *S*s were required to depress the ipsilateral one of two telegraph keys placed 2 cm. right and left of a peg on which *S* rested his right index finger. Instructions were similar to Exp. I and II, stressing a low error rate. Simple RT *S*s used only one of the keys (half the right, half the left) with the unused key covered. Because of a reduction in room illumination, the visual and auditory stimuli were rematched by 10 independent *S*s with a resultant sound level of 64 db. Otherwise, apparatus and procedure were identical to Exp. I and II.

*Results.*—The median warm-up RT for each *S* was entered into one analysis of variance for the simple RT groups or another for the localization RT groups. In both analyses, the groups source of variance was insignificant, $F$s (1, 16) = 1.26 and < 1. This indicated that the groups were equivalent in initial response speed and justified the subsequent between-group analysis of the influence of the mode-predictability variable.

For the analysis of experimental trials, means for each *S*'s 32 trials to the light and 32 trials to the sound were subjected to analysis of variance. Group means and standard deviations are found in Table 3. Standard error of a mean was 5 msec. Mean RT was shorter to the auditory than to the visual stimulus (304 vs. 346 msec.), $F$ (1, 32) = 250.22, $p < .01$. Responses were faster to mode-predictable stimuli (304 vs. 346 msec.), $F$ (1, 32) = 12.83, $p < .01$. And, simple RT was less than localization RT (308 vs. 343 msec.), $F$ (1, 32) = 8.80, $p < .01$. Of greater interest were the only two significant interactions. Task interacted with stimulus modality, $F$ (1, 32) = 28.60, $p < .01$. A Scheffé test indicated that responses were faster to the auditory stimulus in both the simple RT and localization tasks. As Table 1 reflects, the modality difference was 56 msec. for simple RT, but only 28 msec. for localization RT. The other significant interaction was between stimulus mode and mode predictability, $F$ (1, 32) = 4.88, $p < .05$. The effect of making stimulus mode predictable was greater for auditory stimuli (328 vs. 280 msec.) than for visual stimuli (364 vs. 328 msec.).

In the parallel analysis of standard deviations, RTs were more variable to auditory than to visual stimuli, $F$ (1, 32) = 6.80, $p < .05$, and to unpredictable than to predictable stimuli, $F$ (1, 32) = 13.18, $p < .01$. However, mode predictability interacted with task, $F$ (1, 32) = 5.94, $p < .05$, and a Scheffé test indicated that whereas simple RTs were less variable with mode predictable (30 vs. 47 msec.), the variability of localization RTs was unaffected by mode predictability (42 vs. 45 msec.). The Task × Mode interaction was of particular interest because of the differential complexity hypothesis. Although this interaction was not a significant source in the analysis of variance, $F$ (1, 32) = 3.47, $p$

$> .05$, a Scheffé test indicated that localization of auditory stimuli was more variable than visual stimuli, $p < .05$.

## DISCUSSION

All three experiments support the general hypothesis that auditory localization is a more complex process than is visual localization. This is shown by two findings. First, the increment in RT required to localize a sound, relative to simple RT, is greater than the increment in RT for localizing a light, relative to its simple RT base line. Second, although there is no difference in the variability of simple RTs to light and sound, auditory localization RT is more variable than visual localization RT. This pattern holds under variation of stimulus azimuth, stimulus intensity, and response mode.

However, comparison of Exp. I and II with Exp. III also suggests that localization is not independent of response mode. This can be seen by comparing the absolute magnitudes of the RT increments necessary for verbalizing stimulus locus with those for moving a finger in the direction of stimulus locus. In Exp. I and II, using the verbal response, auditory localization added about 40 msec. more to auditory simple RT than visual localization added to visual simple RT. If this value were a "pure" measure of differential perceptual processing, independent of the response mode used to indicate that processing, then the same value should also appear in the manual RTs. But Exp. III, using the manual response, produced a value of 28 msec., indicating that a strict additivity assumption (see LaBerge, 1971) about these data is unwarranted. Another way of seeing this is by considering the facts that verbal localization responses are equally fast to auditory and visual stimuli but manual localization responses are significantly faster to auditory then to visual stimuli. The difference between verbal and manual indicators of localization seems congruent with Simon's (1969) studies. Taken together, the present study and Simon's indicate that pointing is a method of indicating stimulus locus for which the human observer is particularly well atuned. Nonetheless, in spite of this apparent difference in the two methods of response, the relatively greater complexity of auditory over visual localization is indicated by both.

The results of Exp. III indicate that localization processing, per se, does not enjoy flexibility as a function of stimulus mode uncertainty. That is, making the mode of a RT stimulus known to *S* ahead of time produced shorter RT, but the change was the same for both detection and localization conditions. The advantage of mode predictability was greater for auditory stimuli than for visual, although this difference was not large. The analysis of response variability gave further support to the idea that the effect of stimulus mode uncertainty was upon detection rather than localization.

The present data on response latency fit well with earlier findings on response accuracy by indicating that vision is a superior mode for localization. This makes more feasible the possibility (Warren's, 1970, hypothesis) that *S*s may develop the use of some form of visual reference system in performing auditory localization tasks. A more direct test of Warren's hypothesis would now seem to be indicated.

## REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

GELDARD, F. A. *The human senses.* New York: Wiley, 1953.

KOHFELD, D. L. Simple reaction time as a function of stimulus intensity in decibels of light and sound. *Journal of Experimental Psychology*, 1971, **88**, 251–257.

LABERGE, D. On the processing of simple visual and auditory stimuli at distinct levels. *Perception & Psychophysics*, 1971, **9**, 331–334.

PICK, H. L., JR., PICK, A. D., & KLEIN, R. E. Perceptual integration in children. In L. P. Lipsett & C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 3. New York: Academic Press, 1967.

REKOSH, J. H., & FREEDMAN, S. J. Errors in auditory direction-finding after compensation for visual re-arrangement. *Perception & Psychophysics*, 1967, **2**, 466–468.

SIMON, J. R. Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 1969, **81**, 174–176.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

WARREN, D. H. Intermodality interactions in spatial localization. *Cognitive Psychology*, 1970, **1**, 114–133.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental Psychology.* New York: Holt, Rinehart & Winston, 1954.

(Received August 26, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 176–180

# PERCEPTUAL CONDITIONS AFFECTING EASE OF ASSOCIATION[1]

PETER G. ARNOLD AND GORDON H. BOWER[2]

*Stanford University*

Four experiments replicated Asch's findings in 1969 that perceiving two terms in a unitary relationship facilitates their association. In Exp. I and II, pronouncing a CV-C pair as a monosyllable produced high associative recall of the terminal C to the CV cue, while pronouncing the CV-C as two separate terms caused relatively low associative recall. The result held for both intentional and incidental learning conditions. In Exp. III and IV, *S*s were to associate a color patch to each of 15 geometric figures. Association was high when the figure was the same color as the color patch, low when the figure was a different color than the color patch, and intermediate when the figure was an uncolored, black-on-white outline.

Asch (1969) has emphasized the importance of the relationship between perception and memory. In particular, the perceptual relation between pairs of elements may be expected to substantially influence the ease of associating the pairs. Some earlier pertinent data (Asch, 1962) compared the learning of associations between color responses and geometric figures when these were presented in various perceptual relations. The shape and the color could be shown during the study trial in a "constitutive" or "unitary" relationship in which the shape appears to be composed of the colored material. A solidly colored figure on a white background or a colored-outline contour on a white background exemplify constitutive relationships. Alternatively, the shape-color relationship could be nonunitary, as when the figure was a black-outline contour on a colored background or beside a colored patch. The experimental task was to learn to associate color names to a set of black-on-white outline test figures. Learning was about twice as fast when *S*s had studied the shape-color pairings in a unitary than in a nonunitary relationship.

For generality, Asch (1969) performed a similar experiment using verbal materials. Using an incidental learning paradigm, he found that *S*s seeing and pronouncing two CVC trigrams as a unitary disyllable recalled significantly more second syllables when cued with the first than did *S*s seeing the CVCs as a pair of terms and pronouncing them, separated by a pause.

The present experiments attempt to replicate and add further generality to Asch's findings. Experiments I and II investigate the association between a CV bigram and a terminal consonant when these CV-C pairs are presented under unitary (U) and nonunitary (NU) perceptual conditions in an incidental and in an intentional cued recall paradigm. It is expected that the coherence of the CV and C terms in memory will be partly determined by the way they are perceived as being related. Specifically, those *S*s pronouncing the CV-C as a unitary syllable should recall more terminal Cs than those *S*s pronouncing the CV-C pair as two separate units.

Experiments III and IV inquire whether the shape-color results generalize to a paired-associate recall paradigm, in which both colored and uncolored outline figures are paired with a color patch during study trials. (Asch had used an associative matching or recognition test). A novel issue to be investigated is whether the unitary relation between a figure and its surface color can be used to interfere with learning the correct shape-color association. If the surface color of a figure differs from the response color to be associated with

---

[1] This research was supported by a Grant MH-13950 to the second author from the National Institute of Mental Health. Bruno Breitmeyer assisted with Exp. I and III.

[2] Requests for reprints should be sent to Gordon H. Bower, Department of Psychology, Stanford University, Stanford, California 94305.

the shape, then associative interference and response competition should result.

## EXPERIMENT I

### *Method*

Sixteen introductory psychology students, fulfilling a course requirement, were assigned alternately to one of two incidental learning conditions and run in individual sessions. Following completion of the Exp. I task, each of these *S*s participated in Exp. III.

The materials were 15 easily pronounceable CVC trigrams: CUD, DAG, FAS, HIN, KEB, LOM, MEF, NUX, PIX, RIZ, TAV, WEK, WUG, and ZON. The trigrams had meaningfulness scores ranging from 22 to 95, with a mean score of 57 on Archer's (1960) 100-point $a'$ scale. They were lettered, one trigram on each of 15 3 × 5 in. study cards. The initial consonant and the vowel were lettered as a CV bigram approximately .5 in. to the left of center; the terminal consonant, 1 in. to the right of the CV bigram. A test deck comprised 15 cards, each showing just the initial CV.

The *S*s were misinformed that the study concerned the effects of phonetic context upon articulatory voicing. In the NU condition, *S*s were instructed to pronounce the initial CV, then separately to pronounce the final C. They were to pronounce each pair in this manner three times in succession (e.g., to pronounce LO-M as "low-emm, low-emm, low-emm"). In the U condition, *S*s were instructed to blend the initial CV with the final C before pronouncing the CVC as a single unit three times in succession (e.g., to pronounce LO-M as "lom, lom, lom"). The pronunciations of each *S* were tape recorded to contribute to the credibility of the stated purpose of the experiment.

The study cards were presented in random order at a 5-sec. rate. Immediately following the single study trial, *S* was cued with each initial CV for recall of the associated terminal C. If "LO," for example, appeared on a test card, *S* was to respond with "M." Although the recall test was rationalized as an indicator of "the natural coherence of different phonetic compounds," instructions were explicit that *S* was to recall the final C associated with each CV. No *S* reported having anticipated such a recall test. Test cards were presented at a 5-sec. rate in scrambled order.

### *Results*

The recall data supported the experimental hypothesis. When cued with the initial CVs, U *S*s recalled a mean of 7.75 terminal Cs out of 15; NU *S*s, a mean of 4.13. This difference is significant, $t\ (14) = 3.37$, $p < .01$. It may be noted that the similarity between the training stimulus and the test stimulus is greater for the NU *S*s; yet this factor was not as powerful in determining recall as were the different rehearsal procedures.

We next examined the relationship between the recallability of a CVC and preexperimental habits surrounding it. First of all, the CVCs were all selected to be high in pronounceability, so variation in their recall can not be correlated with that. Second, recall of a CV-C pair was correlated with the probability of the final C as a free associate to the initial CV, using Table F from Underwood and Schulz (1960). The rank-order correlations for Groups U and NU and the combined scores were .10 or less and insignificant. Third, recall was correlated with the meaningfulness of the CVCs ($a'$) as scaled by Archer (1960). For the NU *S*s, recallability and meaningfulness did not correlate significantly (rank-order $\rho = .31$). Recallability and meaningfulness did correlate for the U *S*s ($\rho = .5$, $p < .05$) as well as for the combined data of the NU and U *S*s ($\rho = .49$, $p < .05$). The pattern of these correlations suggests that recall differences between the NU and U conditions may be accounted for, in part at least, by differences in the meaningfulness of the phonetic sequences for the two groups. That is, pronunciation of the unitary CVC may remind *S* of a word that it resembles or suggests, whereas the broken CV-C pronunciation may be less likely to do so. This speculative interpretation may not apply, however, to Asch's experiment since he used CVC disyllables.

Experiment II is similar to Exp. I except that it uses intentional instead of incidental learning instructions.

## EXPERIMENT II

### *Method*

Sixteen introductory psychology students, fulfilling a course requirement, were assigned randomly to one of two intentional learning conditions. Eight *S*s, 4 from each of the two conditions, completed the Exp. II task followed immediately by the paired-associate learning task constituting Exp. IV. The remaining 8 *S*s completed the Exp. IV task prior to participating in the present experiment. This counterbalancing was done to control for the possible effect of participation in one experiment on performance in the other. The two conditions were equated for sex and experimental naïveté of the *S*s.

The CVCs, study decks, and test decks used in this experiment were identical to those described in Exp. I. Instructions to *S*s in the present experiment differed from those given in Exp. I in that *S*s were informed that they were to be tested for their memory of the CVCs. This was done to generalize the findings of Exp. I beyond the incidental learning paradigm. All other procedural details were the same as for Exp. I.

### *Results*

The intentional cued recall data were totally consistent with the results of Exp. I. When cued with the initial CVs, U *S*s recalled a mean of 9.62 terminal Cs out of 15; NU *S*s, a mean of 6.12. This difference is statistically reliable, $t$ (14) = 2.57, $p < .025$. The effect was, furthermore, independent of the order in which *S* participated in Exp. II and IV.

Again, recallability correlated somewhat with meaningfulness for the NU *S*s ($r = .45$). The rank correlations for U *S*s and for the combined NU and U data were larger, at .59 ($p < .025$) and .64 ($p < .01$), respectively. Thus, these data are consistent with the differential meaningfulness of phonetic sequences as a possible explanation for the NU versus U recall differences.

## EXPERIMENT III

### *Method*

The 16 *S*s described in Exp. I participated in the present experiment immediately following completion of the Exp. I task. The present experiment was a one-way within-*S*s design.

Fifteen highly discriminable, abstract geometric shapes were randomly divided into three sets of 5. For example, one resembled the outline of a geometrically structured ice cream cone. Within each of the three sets, one of five color response terms was associated with each stimulus. Each shape was drawn on the left side of a 3 × 5 in. card. On the right side of these cards a response color indicator was colored in a 1-sq.-in. patch. Each set of five stimuli was assigned to either the facilitation (F), neutral (N), or interference (I) condition. Shapes of F pairs were colored solidly in the same color as the to-be-associated color patch: shapes of I pairs, in a color different from the to-be-associated color patch. Shapes of N pairs were uncolored figures drawn as black-on-white contours. Three different decks were composed, counterbalancing the assignment of particular sets across the three conditions. Each study deck was used for an equal number of *S*s. There was only one test deck. On each card of this deck was drawn an uncolored black-on-white contour figure identical to that on a study card, except for the missing color. No color patches appeared on the test cards.

Two training trials were given by the study-test method of paired-associate learning with materials presented visually at a 3-sec. rate. No feedback was given on test trials. Study and test items were presented in a different scrambled order each trial. The *S*s were instructed to associate the color patch on the right and the shape of the figure on the left, regardless of the coloring (if any) of the shape itself. A prompting card listing the five color names was available to *S* at all times, and he was encouraged to guess when uncertain.

### *Results*

The mean correct responses per item per *S* over the two test trials were 1.61, 1.30, and 1.15 for the F, N, and I conditions, respectively. An analysis of variance yielded a significant effect of pair type, ($p < .01$). A Newman-Keuls test indicated that more F pairs were correctly recalled than either N or I pairs, but the latter two means did not differ reliably from one another, $q$ (2, 16) = 1.75, $p > .05$. Thus when the unitary relation between a stimulus figure and its surface color was the same as the to-be-learned shape-color association, learning was facilitated relative to the neutral pairs.

Although I items were not recalled significantly worse than N items, *S*s frequently gave as their recall to I shapes the surface color of the shape rather than the appropriate color of the response patch. The frequency of these intrusion errors (IEs) in the I condition suggests that the unitary relation between a shape and its surface color may have been interfering with the acquisition of the correct shape-color association. The IEs accounted for 65% and 45% of all errors in the I condition on Trials 1 and 2, respectively. The probability of this many IEs by chance guessing is remote, $\chi^2$ (1) = 39, $p < .01$.

Assuming that IEs beyond the 25% expected by chance would have been correct responses in the F condition, we may investigate whether the recall of F items is predictable from the recall of I items plus the excess percentage of IEs. Adding percent recall of I items to the excess percentage IEs yields predicted recall scores for the F condition of 67% and 74% for Trials 1 and 2. The comparable observed percentages of F items recalled were 70% and 91%. The prediction for Trial 1 is close, but the Trial-2 prediction is far too low. There thus appears to be no simple quantitative relation between the F and I recall scores.

Asch, Ceraso, and Heimer (1960) found that the association between two geometric shapes (e.g., a rhombus and small crosses) was learned more easily when the elements were related constitutively ("unitary") than when they were contiguous but nonunitary. If a rhombus were outlined by a series of small crosses, for example, the association between the rhombus and crosses was acquired more easily than if the rhombus were outlined by straight lines, with a row of crosses displayed to the right of the shape. Keppel (1962) argued that these results could be explained by the difference in the number of areas in space between which *S* must divide his attention. That is, while *S*s in the contiguous condition had to divide their attention between two fixation areas, *S*s in the constitutive condition had to fixate only one configuration in order to learn the association. The present experiment circumvents that criticism since all study cards had two configurations, the figure and the color patch. The superior recall of F pairs relative to the N pairs despite equated fixation areas challenges Keppel's "twoness" explanation of the earlier results.

## EXPERIMENT IV

### *Method*

Half of the 16 *S*s described in Exp. II participated first in Exp. II and then in Exp. IV. The remaining 8 *S*s participated in this Exp. IV before Exp. II. The design of the present experiment is identical to that of Exp. III, but with a minor change in materials.

The geometric figures, color responses, study decks, and test decks used in the present experiment differed in one respect from those described in Exp. III. For the N and I items, the area of the response color patch was 2 sq. in. rather than 1 sq. in. This change was made in order to equate the total exposure area of the response color patch in the N and I conditions to the sum of the two areas of exposure of the response color in the F condition. This controls for the possibility that previous differences were due to the total area occupied by the correct color on F cards. All instructions and procedural details were identical to those in Exp. III.

### *Results*

The mean correct responses per item per *S* over the two test trials were 1.25, 1.04, and .70 for the F, N, and I pairs, respectively. Analysis of variance revealed significant differences, $F(2, 30) = 14.65$, $p < .01$. Newman-Keuls tests indicated that all three means differed significantly from one another. That is, recall was significantly less for I items than for either F or N items ($p < .01$). The F items, moreover, were recalled reliably more often than the N items ($p < .05$).

Thus the facilitation of recall on the F items relative to N items found in Exp. III was replicated. The Exp. III data had suggested, whereas the present data show more clearly, that correct recall is reduced when the unitary relation between a figure and its surface color conflicts with the to-be-learned shape-color association.

For these I items, *S*s again made a high proportion of intrusion errors. IEs accounted for 53.6% and 43.8% of all errors in the I condition on Trials 1 and 2, which exceeds the chance expectation of 25%, $\chi^2(1) = 32.2$, $p < .001$. Thus the conflicting relations between an I shape and the two colors interfered with learning the correct response.

An attempt was again made to predict percent cued recall of F items from the percent recall of I items plus the excess of IEs. The predicted percent recalls for F items on Trials 1 and 2 were 50.0% and 51.2%. The comparable observed values were 51.2% and 73.8%. Again, while the Trial 1 prediction is close, the Trial 2 prediction is too low.

From these experiments, it may be concluded that the associative effect of perceptual unity is quite robust despite variation in experimental details. Unitary pronunciation facilitates recall relative to nonunitary pronunciation under intentional as well as incidental learning conditions, although the effect may be due to differential meaningfulness of the unified versus separated phonetic compounds. A constitutive relation between a shape and its surface color fosters their natural coherence in memory in comparison to having the color patch simply contiguous to the outlined shape. This natural coherence can be arranged to facilitate or conflict with the to-be-recalled shape-color association.

## REFERENCES

ARCHER, E. J. A reevaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

ASCH, S. E. A problem in the theory of associations. *Psychologische Beitrage*, 1962, **6**, 553–563.

ASCH, S. E. A reformulation of the problem of associations. *American Psychologist*, 1969, **24**, 92–102.

ASCH, S. E., CERASO, J., & HEIMER, W. Perceptual conditions of association. *Psychological Monographs*, 1960, **74** (3, Whole No. 490).

KEPPEL, G. Perceptual conditions of association: A possible confounding. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 214–217.

UNDERWOOD, B. G., & SCHULZ, R. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received August 30, 1971)

---

## ERRATUM

In the article "Modes of Representation and Problem Solving" by Steven H. Schwartz in the November 1971 issue, the question numbered 9 in the second sample problem, column 2, page 347, should be corrected as follows:

The Japanese doesn't live in the blue house and the Indian doesn't drink coffee.

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 181–187

# LIST LENGTH AND METHOD OF PRESENTATION IN VERBAL DISCRIMINATION LEARNING WITH FURTHER EVIDENCE ON RETROACTION[1]

BENTON J. UNDERWOOD, JOHN J. SHAUGHNESSY, AND JOEL ZIMMERMAN[2]

*Northwestern University*

Two list lengths (15 and 45 pairs) and two methods of presenting verbal discrimination lists, study test (ST) and anticipation (A), were used in three trials of original learning. Interpolated learning was either three or nine trials in which the right and wrong functions of the words in original learning were reversed. In five relearning trials, the functions were the same as on original learning. Method had no influence on short lists, but the ST method was distinctly superior to A for long lists. The results of the ST method were judged to be in accord with frequency theory (length is irrelevant); those for the A method were judged not to be in accord. The differences between the methods for the long lists appear to be due to a performance-learning interference. Both interpolated learning and relearning were characterized by initial high performance followed by some deterioration. The failure of the deterioration to be severe even with the long lists poses an explanatory problem. It may be handled by frequency theory only by assuming very high specificity of frequency information.

This experiment focuses on three issues in verbal discrimination (VD) learning: (*a*) the effect of list length, (*b*) the effect of the study-test method as opposed to the anticipation procedure, and (*c*) transfer and retroactive effects when the interpolated VD task consists of a reversal of the right and wrong functions of the members of the pairs. The background for each of these issues will be discussed in turn.

The frequency theory of VD learning holds that length of list will have no influence on rate of learning, but it has also been recognized that completely appropriate tests are difficult to implement (Ekstrand, Wallace, & Underwood, 1966). If words are used to form the lists, interitem associations among them may disturb the frequency relationships between the right (R) and wrong (W) items within a pair, and it is the intrapair frequency relationship which is critical in making predictions. As list length increases the likelihood of the presence of interitem associations among the words also increases. Unless such associations could be equated for long and short lists, the test of the theory would not be considered appropriate. For various reasons, it seems unlikely that such equation could be accomplished. In the present study, a different approach has been taken. Low-frequency words have been used under the assumption that differential effects of interitem associations will be minimal for different list lengths. Likewise, other factors which might influence learning differentially as length increases, factors not identified by the theory, are also assumed to have minor effects on learning. Therefore, the theory must be prepared to accept small differences as a function of list length. The problem of just what may be considered "small" can be met after the results are examined.

The second variable in the experiment is the method of presentation, study test (ST) versus anticipation (A). At least three studies have led to the conclusion that VD learning occurs more rapidly with the ST method than with the A method (Battig & Switalski, 1966; Fulkerson & Johnson, 1971; Ingison & Ekstrand, 1970). As Battig and Switalski pointed out, the results run contrary to the notion that immediate feedback is favorable for learning. It is their belief that the ST method clearly separates the learning and performance functions required of *S*, whereas the A method does not, but just how this leads to performance differences under the two methods is not clear. If delay of feedback is to have any influence on VD learning by the ST method, the effect should appear with long lists since list length determines the delay. It is possible, therefore, that with long lists the A method might become superior to the ST method. On the other hand, if the separation of learning and performance is the critical factor, length of list should be irrelevant. In any event, further study of this variable seems necessary, both to identify more precisely the source of the difference produced by the two methods and to discover if it represents the operation of a process that cannot be handled by frequency theory as now formulated.

---

[1] This work was supported by Contract N00014-67-A-0356-0010, Project 154-057, between Northwestern University and the United States Office of Naval Research. Reproduction in whole or in part is permitted for any purpose by the United States Government.

[2] Requests for reprints should be sent to Benton J. Underwood, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

If, following learning of a VD list, the functions of the members of each pair are reversed (the R becoming W, and W becoming R), performance is initially high but error reduction occurs at a relatively slow rate as trials continue (Underwood & Freund, 1970). Following learning of the interpolated list to a high level, recall of the original list is very high but again subsequent relearning is impaired. Both the initial high level and the subsequent deterioration are quite in line with frequency theory except for the quantitative aspects of the deterioration. Strictly speaking, the theory would say that performance should fall to a near chance level and it doesn't. One possibility is that other discriminative attributes replace frequency when the frequency discrimination breaks down. If this is so, it would seem that replacement would be far more difficult with a long list than with a short list. Therefore, the third purpose of the experiment is to discover if the deterioration in performance during interpolated learning and during relearning of the original list will be greater for a long list than for a short list.

The general design may now be summarized. Two list lengths (15 and 45 pairs) and two methods of presentation (ST and A) were employed with a given *S* having a constant list length and method during all three stages (original learning, interpolated learning, and relearning). Half of the *S*s had a low level of interpolated learning, half a high level.

## METHOD

*Materials.*—A total of 90 two-syllable words was used to form the pairs. These 90 words were chosen randomly from a pool of 832 words which had in turn been drawn randomly from the Thorndike-Lorge tables of words having frequencies of from 1–10 in the G count. The 90 words were paired randomly and the pairs formed into three 15-pair lists. These 15-pair lists formed the basic units for the balancing to be described. For the conditions requiring 15-item lists, each of the three lists was used equally often across *S*s. The 45-item lists were formed by combining the three 15-item lists. If the three lists are labeled A, B, and C, three different 45-item lists were formed by the orders ABC, BCA, and CAB. These three lists were used equally often. It should be noted that as implied by the units (ABC), the orders of the 45-item lists across trials were not changed by randomizing across all 45 pairs. Rather, the first 15 pairs remained the same across trials for a given *S* just as did Pairs 16–30 and Pairs 31–45. However, within these three blocks, pairs were randomized from trial to trial to form six different orders. This scheme makes it possible to compare performance on the first 15 pairs in the long list with the same pairs in the short list.

*Design and procedure.*—There were eight groups formed by two list lengths (15 and 45 pairs), two methods (ST and A), and two levels of interpolated learning (three and nine trials). By block randomization, 27 college-student *S*s were assigned to each condition. Further included in the randomization was the particular list assigned *S*, there being 9 *S*s allotted to each of the three different lists under both the 15- and 45-pair conditions.

The original learning (OL) consisted of three performance trials under all conditions. For both ST and A, *S*s were first given a study trial which was exactly the same for both methods. On this study trial, the two words of a pair were shown, one above the other, on the left side of the memory-drum tape, with the correct member of the pair being printed to the right. All three terms were present simultaneously for 2 sec. After the study trial, the ST and A procedures were instituted. Under A, the shutter was activated and the timing changed to produce a 2:2-sec. rate. Three anticipation trials were given as OL. Under the ST method, a test trial was given immediately after the initial study trial. On the test trial, each pair was shown at a 2-sec. rate and *S* was required to choose the correct member of each pair. After the test trial, a study trial was given which was exactly the same as the initial study trial (except, of course, for the order of the pairs). Three test-study cycles followed the initial study trial. For these *S*s, therefore, the last trial of OL consisted of a study trial. It should be noted that the total time involved in the two methods was exactly the same. The only difference was that the A method included both study and test on three trials (4 sec. per pair), while the ST method separated the study and test into three trials of each with a 2-sec. rate per pair. The intertrial interval under ST was 2 sec., under A, 4 sec.

Prior to OL, *S*s were fully instructed concerning the task and the method to be followed. Immediately following OL, *S* was instructed concerning the nature of interpolated learning (IL), with particular emphasis on the reversal of functions for the R and W items. The instructions required *S* to respond on the first presentation of the list in IL. Following IL, the same explicit instructions were given concerning recall and relearning of the first list and again *S* was asked to respond on the first presentation of the list. The IL for the ST method always ended with a study trial. Thus, with nine IL trials there were nine test-study cycles, and with the A method, nine anticipation trials. Relearning was carried for five anticipation trials or five test-study cycles.

It should be noted that the time between OL and relearning differed as a function of degree of IL and as a function of list length. The difference ranges from 3 min. (15-pair lists, three IL trials) to 27 min. (45-pair lists, nine IL trials). We justify this confounding on the basis of previous work (Underwood & Freund, 1970), which shows that the forgetting of a VD list is extremely slow (9% loss over 24 hr. following three trials on a 42-pair list).

## RESULTS

*Original learning.*—Performance on the three OL trials can be used to assess the effects of method and list length on VD learning. By combining groups differing only on degree of IL, 54 *S*s were available for each unique condition. The performance curves, expressing percent correct on each trial, are shown in Fig. 1. Performance on the 15-pair lists under the two methods was identical on the second and third trials and differed by 1% on the first; therefore, a single line is used to represent these two groups (ST-A-15). It is apparent from Fig. 1 that the two variables of interest interact: with the 15-pair lists method is of no consequence, whereas with the 45-pair lists the ST method results in a higher acquisition level than does the A method. The statistical analysis (following an arc-sine transformation) supports these observations. With trials, list forms, and degree of IL included in the analysis, the *F*s for length (74.72), method (13.83), and Length × Method (10.31) were all significant beyond the 1% level (1, 192 *df*). The only other significant sources of variance were trials and two third-order interactions (5% level) involving list forms and degree of IL which are not given to interpretation.

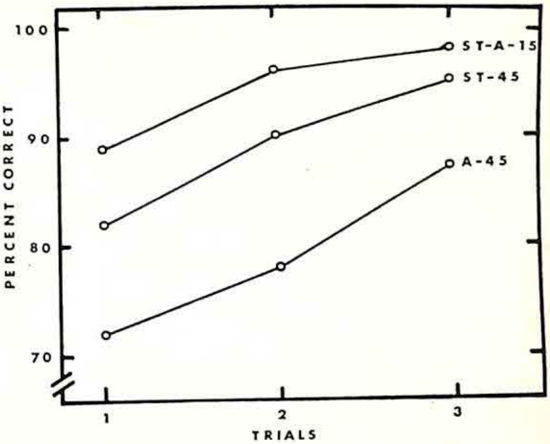


FIG. 1. Learning as related to method of presentation (ST and A) and length of list (15 and 45 pairs).

For the 45-pair lists, for the first trial, performance is better for the ST method than for the A method. This is true in spite of the fact that the initial study trial was exactly the same for both methods; hence, learning should have been equivalent. This strongly suggests that under the A method the ability to perform what has been learned is inhibited or interfered with by the new learning which is occurring. It will be remembered that the 45-pair lists consisted of three blocks of 15 pairs each which, as blocks, maintained a constant order across trials. Thus, the retention interval from trial to trial was constant for the blocks (although as noted earlier this should be of little if any consequence). An examination of the performance on the three successive blocks revealed decreasing performance across the three blocks for both methods, the difference between the first and third blocks

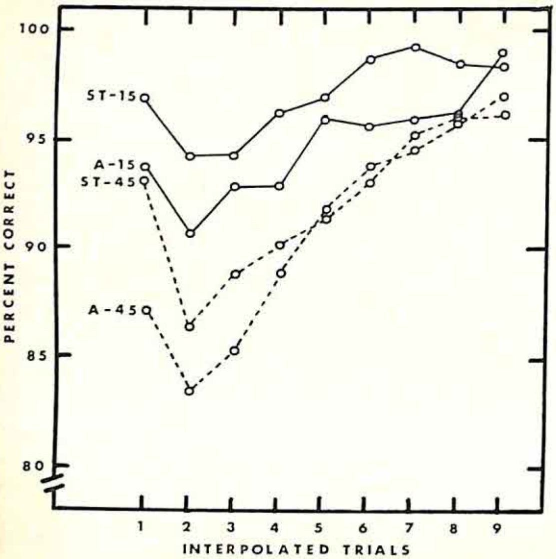


FIG. 2. Interpolated learning as related to method of presentation and length of list.

being approximately one item for both methods. For the ST method, this position effect was not present on Trials 2 and 3; for the A method, it remained on the second trial but not on the third. No position effects were present in the 15-pair lists. An examination was made of performance using the first 15 pairs only. As would be apparent from the above facts, this reduced the difference between the short and long lists. However, across the three trials the difference between the long list learned under the ST method and the corresponding short list would still be judged to be significantly different, $F(1, 107) = 4.10$, $p < .05$. In summary, method differences occur only for the long lists, and length is a statistically significant variable, but the influence of length is greater under the A method than under the ST method.

*Interpolated learning.*—The performance curves for the four groups having nine IL trials are shown in Fig. 2. Differences in levels must be viewed with caution because of differences in terminal level of OL, differences which cannot be specified exactly because the learning which occurred on the last trial was not measured. But, as can be seen, performance level is roughly correlated with degree of learning in OL. The differences between the two short lists (although not significant statistically) also may reflect a difference between these two groups which was present in OL. All curves show the decrement expected by frequency theory; i.e., performance on the first trial is better than performance on the immediately succeeding trials. The amount of the fall is not large in an absolute sense and statistically does not differ for the four conditions. To make this analysis, a score was determined for each *S* representing the difference between his performance on the first trial and the lowest point of performance on the succeeding four trials. As would be expected from Fig. 3, the lowest point of performance for most *S*s was on the second IL trial. The percentage fall (or in a few cases a gain) was determined for each *S*, and after an arc-sine transformation the analysis of variance was performed on the four distributions. No source of variance was significant.

It was noted above that in an absolute sense the decrement in performance after the first trial was not great. Most errors occurred on items on which there had been errors during OL. For the 15-pair lists approximately 81% of the pairs were correct on all OL trials, and for the 45-pair lists the corresponding value was 66% (combined across methods). In IL, these pairs were responded to correctly on all trials from 77% to 87% of the time for the various conditions. In short, the typical pair is one which was always correct in OL and always correct in IL.

*Recall and relearning.*—Method (ST vs. A) was not a significant source of variance in recall and relearning, so the data have been combined across methods for presentation in Fig. 3. For both lists length recall was higher following nine IL than following three IL, but with nine IL performance falls on subsequent trials so that on the fifth relearning trial the scores are a little higher following three IL trials than following nine. For the nine IL groups, performance never achieves the level exhibited on the recall trial. All condition differences (15 vs. 45 pairs; three vs. nine IL trials) are significant statistically, as are the first-order interactions with trials so apparent in Fig. 3. Finally, the Length × Degree of IL × Trials interaction, also apparent in Fig. 3, is also significant statistically, $F\ (4,\ 768) = 17.15,\ p < .01$. This indicates that the differences over trials between three and nine IL trials for the 15-pair lists is less than the corresponding difference for the 45-pair lists.

Overall, the relearning data indicate that with a high degree of IL, reversal of item function (to correspond to OL) is almost perfect initially, but immediately thereafter interference occurs so that very little increase in performance occurs over trials. When OL and IL are essentially equal in level of learning (as with three IL trials), reversal of function is initially poor, relative to that following a high level of IL, but improves rapidly with additional trials. While these facts are clear, it should again be pointed out that in an absolute sense the changes in performance as described are not large and for all groups performance is at a high level over all five trials.

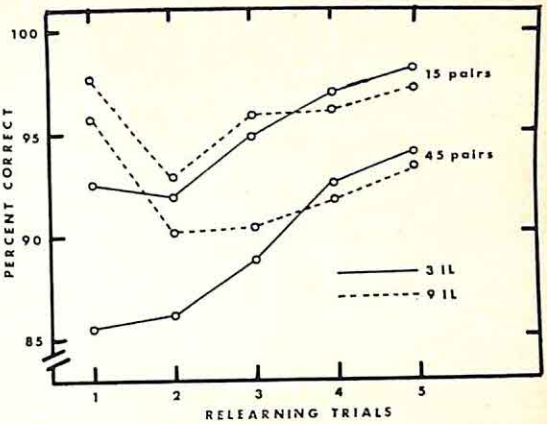


FIG. 3. Relearning as related to degree of interpolated learning and length of list.

## DISCUSSION

Three issues will be discussed: the method differences, the length differences, and the transfer and retroaction data as they reflect on the frequency theory of VD learning.

The results have shown that with long lists the ST method produces more rapid learning than does the A method. This difference occurs on the initial part of the first performance trial even though the study trial was exactly the same for both methods. This supports the idea of Battig and Switalski (1966), that when *S* must both exhibit learning which had occurred on the earlier trial and undertake new learning, as seems necessary by the anticipation method, performing the learned discrimination is inhibited. For the 45-pair list, the differences between ST and A amounts to about 5 pairs on the first trial; that is, 5 more pairs are given correctly under the ST method than under the A method, although learning must have been equivalent. By frequency theory, once those 5 pairs are incorrectly responded to, the discrimination which had been established on the study trial would be lost and these items will not be benefited appreciably on the immediately succeeding trials by the original discrimination. The 15-pair lists were not differentially influenced by the two methods. This is contrary to the results of three other investigations reported in the introductory section, where the number of pairs varied from 12 to 20. The reason for the contradictory findings are not known. However, referring now only to the present results, a hypothesis will be advanced to account for the Length of List × Method interaction. This hypothesis holds that the magnitude of the difference between the two methods will be directly related to the number of unlearned discriminations that exist after the initial study trial (or on any subsequent trial). The short lists in the present study resulted in slightly over 13 of the 15 pairs being correctly responded to after the initial study trial, and it must be remembered that in VD learning the probability of being correct on the second trial for items correct on the first is very high (.97 for the short lists in the present study). For the long lists, at least seven items remained unlearned after the first study trial. As learning proceeded, however, the number of unlearned items diminished and, as shown in the interpolated learning for the long lists (Fig. 2), performance became equivalent under the two methods. The hypothesis would anticipate differences in performance under the two methods for short lists if, for whatever reason, the learning on the initial trial is low. This might be accomplished, for example, by using nonintegrated units such as nonsense syllables.

The next issue concerns list length per se. As pointed out in the introduction, frequency theory can accept a small effect of list length on the grounds that it is essentially impossible to equate all factors except length as list length increases. The question is whether or not the effects of list length as found in the present study are of a magnitude that will allow them to be classed as "small." To arrive at a reasonable assessment it seems necessary to relate the results to those found in other tasks where associative learning is involved. The evidence available indicates that for associative learning, as length of list increases the *percentage* of items learned decreases *markedly*. For example, in a study by Postman and Phillips (1956), Ss were given a single study trial on unrelated lists of words, with the lengths of lists being 10, 20, and 30. Each word was presented for 1 sec., after which *S* recalled all the words possible. The recall, as measured by percents was 56%, 31%, and 26% for the three lengths of lists in order. It is estimated that a 15-item list would produce 45% recall, a 45-item list, 20%. Now, it is apparent from Fig. 1 that the differences in the present VD lists are not of this magnitude. After a single study trial, the 15-pair lists had 89% of the items correct; for the 45-pair lists, the values were 82% for ST and 72% for A. The difference of 7% for the 15- and 45-pair lists learned by the ST method is not large, although it is significant statistically. Even if only the first 15 pairs are scored a difference remains, although on the first trial it is only 5%. Whether this small difference is due to interitem associations or formal similarity which disturbs the intrapair frequency relationships is unknown, but it must be judged to be within the range of "small" effects which frequency theory can tolerate. In the case of the A method, the differences are of appreciably higher magnitude. If the analysis presented earlier is correct, the longer list increases the importance of the performance-learning conflict which decreases performance. Frequency theory has no provisions for handling this effect. Therefore, the theory must state that list length will be of small consequence only when the study-test method is used.

The final issue concerns the evaluation of frequency theory in the light of the transfer and retroaction results. According to the theory, OL is acquired by using the frequency differential between R and W words as the discriminative cue. When moving to IL, *S* simply reverses his rule, now choosing the member of each pair with the lower frequency. The original rule is then applied at the time of recall of the original list when the degree of IL is high. The critical problem for the theory, one that applies both to IL and relearning, is the failure of performance to break down badly when frequency is no longer an appropriate cue for the discrimination, as it clearly cannot be during certain points in IL and relearning. The failure of performance to deteriorate badly was also noted in an earlier study (Underwood & Freund, 1970), and it was suggested at that time that *S* may have switched to other cues for the discriminations. This seemed feasible for a 15-pair list; but it does not seem reasonable that this could be done so expeditiously for a 45-pair list. The present data show, however, that the magnitude of the fall following the initial high performance on IL does not differ as a function of the length of the list. So, the question remains as to how *S* can perform so well in IL and relearning, showing only a relatively small decrement. And, to emphasize a fact, the typical pair is responded to correctly on all OL trials and on all IL trials.

Hintzmann and Block (1971) have shown that frequency information may be quite specific to a given task. They have suggested, furthermore, that this fact might be used to resolve the problem which is the concern of the discussion of the moment, the failure to find the deterioration expected by frequency theory. Their speculation runs about as follows. On the first IL trial, *S* must of necessity respond with information carried over from OL. This may continue into the second IL trial. However, at the same time the frequency inputs are starting from scratch for IL, and even after a single trial the differential in frequency between members of the pairs would be sufficient to allow correct responding a high proportion of the time. Therefore, when frequency information from OL begins to fail as a basis for discrimination, *S* switches to the new frequency information which has accrued for IL. While the high specificity and independence of the frequency information in OL and in IL may seem quite implausible, one is faced with what may be perhaps an even less plausible alternative, namely, that *S* switches the basis for his discriminations and somehow sorts the 90 words (in the 45-pair list) into two categories using, perhaps, semantic attributes as the cues. That this could be done so efficiently seems unlikely, although there is very little evidence on learning of this nature. In a study by Paul (1970), 24 words were presented individually with *S* required to respond in one way for 12 of the words (selected randomly) and in another way for the other 12. Learning did occur, of course, but after 10 trials performance was slighly under 75% correct. To categorize 90 unrelated words into two groups would seem to offer a very difficult task for *S*.

The Hintzman-Block hypothesis cannot be dismissed, which is to say that frequency information can be highly specific to a given time and task. However, the acceptance of this notion must be conditionalized on the acceptance of the further notion that *S* has some control over the counting mechanisms. This is to say frequency information does not necessarily start at a base of zero for each change in task. For example, several studies have shown that pretraining by free recall learning on the R or W items prior to VD learning will influence VD learning substantially. It would appear that *S* can make some judgment of the appropriateness or inappropriateness of the frequency information available and act (to some degree at least) accordingly. It might be maintained, of course, that the problem which has been the focus of the discussion here arises only because of the assumption that the original VD learning is based on frequency information. Yet, if some other theory is used to account for VD learning, particularly a theory based upon associative learning, one critical issue must be met at the outset and that is the relatively small effect of length of list on VD learning.

## REFERENCES

Battig, W. F., & Switalski, R. W. Comparison of anticipation and recall procedures in verbal discrimination learning. *Psychonomic Science*, 1966, **6**, 65–66.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Fulkerson, F. E., & Johnson, J. E. Methodological variables in verbal discrimination learning. *Psychonomic Science*, 1971, **22**, 68–69.

Hintzman, D. L., & Block, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297–306.

Ingison, L. J., & Ekstrand, B. R. Effects of study time, method of presentation, word frequency, and word abstractness on verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **85**, 249–254.

Paul, C. Acquired equivalence of correct alternatives after verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **86**, 123–125.

Postman, L., & Phillips, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 132–138.

Underwood, B. J., & Freund, J. S. Retention of a verbal discrimination. *Journal of Experimental Psychology*, 1970, **84**, 1–14.

(Received September 4, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 188–192

# GENERAL AND SPECIFIC COMPONENTS OF RETROACTIVE INHIBITION IN THE A-B, A-C PARADIGM[1]

ISABEL M. BIRNBAUM[2]

*University of California, Irvine*

Two experiments investigated the effects of first-list response suppression and specific associative unlearning on retroactive inhibition (RI) in the A-B, A-C paradigm. Either two or four List 1 stimuli were omitted from interpolated lists, and two different tests of recall were used. Contrary to prior findings by I. M. Birnbaum in 1970, when stimuli were omitted from the interpolated lists, RI for the corresponding List 1 pairs was significantly reduced. Combined analyses of prior and present results demonstrated that specific associative loss as well as first-list response suppression makes a significant contribution to RI in the A-B, A-C paradigm.

The present paper is concerned with the contribution of specific and general components of interference to retroactive inhibition (RI) in the A-B, A-C paradigm. According to the assumptions of interference theory (Keppel, 1968), specific associative unlearning occurs when a particular List 1 stimulus term appears during interpolated learning (IL) and elicits the now incorrect List 1 response term. Postman, Stark, and Fraser (1968) suggested that suppression of the first-list response repertoire rather than specific associative loss may constitute the major source of interference in the A-B, A-C paradigm. Postman et al. assumed that "specific associations are in general highly resistant to unlearning [p. 689]." If RI is attributable solely to the suppression of the entire first-list response repertoire, then the omission of a subset of List 1 stimuli from the interpolated lists should not eliminate or substantially reduce the amount of RI for the corresponding List 1 pairs. Confirmation of this deduction was reported recently (Birnbaum, 1970). The omission of either two or four List 1 stimuli from one or two interpolated A-C lists did not significantly reduce the amount of RI for the corresponding List 1 pairs.

The foregoing study used a retention test which was assumed to maximize first-list response suppression; that is, the interpolated responses were made available to *S*s at the time of the test. Postman et al. (1968) observed that this procedure tended to maintain suppression of the first-list response repertoire. The impetus for the present set of studies was to investigate the extent of first-list response suppression when *S*s were asked to recall all responses to each stimulus (the typical modified-modified free recall (MMFR) test). As will be seen, the search for the relationship between first-list response suppression and conditions of testing revealed that there is a significant contribution of specific associative loss to RI in the A-B, A-C paradigm under either test condition.

## EXPERIMENT I

### *Method*

*Design.*—Two experimental groups learned a total of three paired-associate lists which conformed to the A-B, A-C, A-D paradigm. A Control group learned only the first list and engaged in a numerical reasoning task during the retention interval. For the experimental groups, two of the 12 List 1 stimuli were omitted from the second and third lists. Thus, *S*s learned a 12-pair first list and two 10-pair interpolated lists. Over the entire experiment, six different sets of 2 stimuli were used as omitted stimuli, and each set was used equally often in each group. The List 1 pairs from which stimuli were omitted on the interpolated lists will be designed OM pairs; the List 1 pairs from which stimuli were present on the interpolated lists will be designated PR pairs. On the self-paced retention test, *S*s in the First-List Recall group (FLR) were

---

[1] This research was supported by United States Public Health Service Research Grant MH-13629. Thanks are due to Al Ahumada, Jr., and James F. Voss for their valuable comments and suggestions at various stages of the research.

[2] Requests for reprints should be sent to Isabel M. Birnbaum, School of Social Sciences, University of California, Irvine, California 92664.

provided with the entire set of 20 interpolated responses and were asked to recall the first-list responses. The *S*s in the Total Recall group (TR) were asked to recall all of the responses which had been paired with each stimulus. After the data from Exp. I were analyzed, a complete replication of Group FLR was run.

*Materials.*—Three lists of 12 CVC-adjective pairs were used equally often in each group for original learning (OL). Two of the lists had been used in the previous study (Birnbaum, 1970), and the third was constructed according to the same rules. The lists for IL were selected from two additional lists, X and Y, which contained the 12 List 1 stimuli and two sets of 12 new and unrelated adjectives as responses. Six different sets of two 10-pair interpolated lists were derived from Lists X and Y. In constructing the six sets, the 2 pairs to be eliminated from X and Y were selected at random (without replacement), with the restriction that the two omitted stimuli contain a different vowel. For a particular *S*, the same stimuli were omitted from both interpolated lists. The second and third lists learned were derived from List X and List Y, respectively. One of the six sets of interpolated lists was randomly assigned without replacement to each of the six *S*s who learned a particular first list in each group. Three of the four orders of presenting the pairs were used equally often as starting orders.

*Procedure.*—The lists were presented at a 2:2-sec. rate with a 4-sec. intertrial interval. Details of procedure may be found in Birnbaum (1970). In brief, all *S*s were taken to a criterion of 8/12 correct in OL. There was a 4-min. interval between the end of OL and the first study trial in IL. Groups FLR and TR received one study trial and six anticipation trials on each of the interpolated lists. The instructions made no mention of the omission of two stimuli. The interval between the end of OL and the beginning of the recall test was 20.5 min. for all groups. Prior to the recall test, *S*s in Group FLR were given an appropriate index card on which the second-list and third-list responses were printed in alphabetical order. The *S*s were told that the list might help them to avoid confusion between first-list and interpolated-list responses, and they were instructed to recall the first-list response to each stimulus as it appeared in the drum window. Hence, the retention test used in Group FLR was identical to the retention test used for all experimental groups in the previous study (Birnbaum, 1970). Group TR, on the other hand, was instructed to recall all of the responses which had been paired with each stimulus. In all groups, the recall test was paced by *S*, and the 12 stimuli appeared in the order which would have been next in OL. The task for the Control group was to recall the single response that had been paired with each stimulus in OL.

*Subjects.*—The *S*s in Exp. I were 54 male and female undergraduates at the University of California, Irvine. Eight additional *S*s were discarded because of procedural error or apparatus failure. The *S*s were assigned to groups in blocks of 3 in order of their appearance at the laboratory, with 1 *S* per group in each block. The order of running groups within blocks, assignment of OL lists, starting orders, and IL lists were determined by a table of random numbers subject to restrictions necessary for balancing each of these factors within groups. The 18 *S*s who participated in the replication of Group FLR came from the same population and were run in the same academic quarter.

### *Results and Discussion*

*Original learning.*—The mean number of trials to criterion in OL was 7.6, 8.6, and 6.2 in the Control group, Group FLR, and TR, respectively. A two-way analysis of variance of trials to criterion indicated that there was no significant difference between groups, $F\ (2, 45) = 2.06,\ p > .10$. The three OL Lists differed in number of trials to criterion, $F\ (2, 45) = 3.45,\ p < .05$, and there was no significant interaction of Groups × Lists, $F\ (4, 45) = 1.39,\ p > .10$. In all subsequent analyses of first-list recall, there was no significant main effect of OL lists and no significant interaction involving OL lists. This variable will not be mentioned again. The mean number of trials to criterion in the replication of Group FLR was 7.5.

*First-list recall.*—A response was scored correct when it was given to the appropriate stimulus. The mean total number of correct responses was 9.72, 3.44, and 4.39 in the Control group, Group FLR, and Group TR, respectively. A two-way analysis of variance of number correct at recall yielded one significant effect, groups, $F\ (2, 45) = 57.77,\ p < .01$. Clearly, there was a substantial amount of RI in both experimental groups. The difference between Group FLR and Group TR was not significant, $F\ (1, 45) = 2.24,\ p > .10$.

The phenomenon of major interest is the recall of OM and PR pairs in Groups FLR and TR. For the purpose of comparison with the experimental groups, OM versus PR was used as a pseudovariable in scoring recall for the Control group. Two pairs for each *S* in the Control group were designated as "OM pairs" by random assignment and counterbalancing across OL lists, as in the experimental groups. The recall scores for OM and PR pairs

TABLE 1

MEAN SCALED RECALL AND PERCENT RETROACTIVE INHIBITION (RI)

| Type of pair | Group | | |
|---|---|---|---|
| | Control | FLR | TR |
| PR | 1.61 | .54<br>(66%) | .62<br>(61%) |
| OM | 1.66 | .72<br>(57%) | 1.28<br>(23%) |

Note.—FLR = First-List Recall group and TR = Total Recall group. PR = pairs from which stimuli were present on the interpolated lists, and OM = pairs from which the stimuli were omitted.

were made directly comparable by expressing both the OM and PR score for each *S* on a scale from zero to two. Hence, the scaled recall score equals 2 × (number correct/maximum possible number correct). All analyses reported below were done on this measure. The mean scaled recall score for OM and PR pairs is shown in Table 1, with percent RI in parentheses below each entry. The means for OM and PR pairs are virtually identical in the Control group, as would be expected with the complete counterbalancing of specific pairs in combination with OL lists. Relative to the percent RI for PR pairs, there is a considerably greater reduction in percent RI for the OM pairs in Group TR than in Group FLR. A three-way analysis of variance of recall for the experimental groups only (groups, OM versus PR pairs, and OL lists) yielded a significant effect of OM versus PR, $F\ (1, 30) = 10.08$, $p < .01$. For the interaction of Groups × OM versus PR, $F\ (1, 30) = 3.31$, $p < .10$. Separate analyses of variance of recall in each experimental group yielded a significant effect of OM versus PR only in Group TR, $F\ (1, 15) = 12.28$, $p < .01$; for Group FLR, $F\ (1, 15) < 1.00$ for OM versus PR. A two-way analysis of variance of recall of OM pairs alone for the Control group and Group TR yielded $F\ (1, 30) = 5.21$, $p < .05$, for groups. Hence, in Group TR the omission of List 1 stimuli substantially reduced but did not eliminate RI for the OM pairs.

The results indicate that suppression of the entire set of List 1 responses was a strong influence on recall in Group FLR, since OM pairs showed almost as much RI as did PR pairs in this group. In contrast, Group TR showed considerably less RI for OM than for PR pairs. Since both OM and PR pairs were subject to suppression of the first-list response repertoire, while only PR pairs were subject to unlearning during IL, these results suggest that specific associative unlearning was an important source of retention decrement in Group TR. Furthermore, since RI for PR pairs was only slightly greater in Group FLR than in Group TR (66% vs. 61%), it appears that specific and general components of RI are not additive (opportunities for specific associative unlearning during IL were, of course, the same in both groups). Hence, providing *S*s with the interpolated responses at the time of recall (Group FLR) may maintain suppression of the entire set of List 1 responses, but may also mask the influence of specific associative unlearning.

It seemed advisable to replicate Groups FLR and TR, but the coming end of the school year precluded a complete replication. The decision was made to run 18 *S*s in Group FLR, and the same *E* conducted the replication. The outcome of the replication was surprising. Mean scaled recall of PR pairs was .64, and mean scaled recall of OM pairs was 1.27. The difference between OM and PR was significant, $F\ (1, 15) = 15.54$, $p < .01$. The replication of Group FLR essentially duplicated the results previously found in Group TR (see Table 1). At this point, it appeared that the type of retention test was not a critical factor in the recall of OM and PR pairs. Instead, the results indicated that OM pairs probably have a true advantage over PR pairs at the time of recall. The advantage is small, however, and recall may be subject to many other influences (e.g., idiosyncratic differences between *S*s in the difficulty of particular List 1 pairs, etc.).

It will be recalled that the previous study (Birnbaum, 1970) showed that the omission of particular List 1 stimuli did *not* reduce the amount of RI for the corresponding List 1 pairs (OM pairs). It now appears that the small and nonsignificant differences observed in that study were attributable to a real difference between OM and PR pairs. The OM versus PR differences were evaluated using a between-*S*s design. Recall of actual OM and PR pairs was compared with recall of "mock" OM and PR pairs. The design of the present study permitted a simple within-*S*s comparison, a more sensitive measure of the effect of OM versus PR. It should be noted that a superiority of recall of OM pairs in the previous study was suggested only in the groups for which two rather than four stimuli had been omitted in IL. Hence, the present study may have used the most favorable procedures for detection of a difference between OM and PR pairs, a within-*S*s design and omission of only two List 1 stimuli from the IL lists. In the next experiment, four List 1 stimuli were omitted from the interpolated lists.

## EXPERIMENT II

### *Method*

*Design.*—Two groups learned three paired-associate lists which conformed to the A-B, A-C, A-D paradigm. Four of the 12 List 1 stimuli were omitted from the second and third lists. Over the entire experiment, three different sets of 4 stimuli were used as omitted stimuli, and each set was used equally often in each group. On the self-paced test of recall of List 1, *S*s in Group FLR were provided with the entire set of 16 interpolated responses, while *S*s in Group TR were asked to recall all of the responses which had been paired with each stimulus.

*Materials.*—The three lists used for OL were the same as those used in Exp. I. For IL, three different sets of two 8-pair interpolated lists were derived from X and Y (Exp. I). The rules for construction of IL lists and assignment of IL lists to *S*s were based on the principles used in Exp. 1. Each List 1 stimulus was omitted from only one of the three sets of IL lists, and each set of IL lists was randomly assigned to two of the six *S*s who learned a particular OL list in each group.

*Procedure.*—Most of the details of procedure were the same in Exp. I and II, with the following differences. Shorter IL lists in Exp. II produced a shorter retention interval, 18.75 min. instead of 20.50 min. Also, the instructions for recall in Group TR stated that "some of the nonsense syllables from the first list also appeared in the second and third lists you learned." In Exp. I, the instructions for recall in Group TR made no mention of whether some or all of the List 1 stimuli had appeared on the IL lists.

TABLE 2

MEAN SCALED RECALL AND ESTIMATED PERCENT RETROACTIVE INHIBITION (RI)

| Type of pair | Group | |
|---|---|---|
| | FLR | TR |
| PR | .82<br>(50%) | .93<br>(43%) |
| OM | 1.11<br>(32%) | 1.22<br>(26%) |

Note.—FLR = First-List Recall group and TR = Total Recall group. PR = pairs from which stimuli were present on the interpolated lists and OM = pairs from which the stimuli were omitted.

*Subjects.*—The *S*s were 37 paid volunteers who were recruited in summer session classes and campus student centers at the University of California, Irvine. One of the *S*s was discarded because of a procedural error. The assignment of *S*s to groups was made in blocks of 2, with 1 *S* per group in each block. The assignment of lists, balancing within groups, etc., followed the rules used in Exp. I.

### *Results*

*Original learning.*—The mean number of trials to the criterion of 8/12 correct in OL was 6.94 in Group FLR and 7.94 in Group TR. A two-way analysis of variance of trials to criterion yielded $F$s $< 1.00$ for groups and the interaction of Groups × Lists and, $F\ (1, 30) = 1.95$, $p > .10$, for lists.

*First-list recall.*—The mean total number of List 1 responses given to the appropriate stimulus was 5.50 and 6.17 in Groups FLR and TR, respectively. A two-way analysis of variance (groups; OL lists) of total number correct at recall yielded no significant main effects or interactions, all $p$s $> .10$.

The mean recall of OM and PR pairs expressed on a scale from zero to two is shown in Table 2. On the basis of the results of Exp. I, it was assumed that the mean scaled recall for OM and PR pairs in a control group would have been approximately 1.64. The estimated percent RI was calculated using this figure and is shown in parenthesis below each entry. A three-way analysis of variance (Groups; OM vs. PR pairs; OL lists) of scaled recall yielded one significant effect, OM versus PR, $F\ (1, 30) = 6.94$, $p < .05$. For all other main effects and interactions, $ps > .10$.

## General Discussion

Comparing the results of Exp. I and II, the absolute level of recall of OM pairs for experimental groups was similar in both experiments and, in general, was significantly higher than the recall of PR pairs. Mean scaled recall of PR pairs was somewhat higher in Exp. II than in Exp. I, as would be expected on the basis of the shorter IL lists used in Exp. II (Birnbaum, 1970). It appears that the unlearning of specific first-list associations made a substantial contribution to the observed retention losses. The PR pairs, subject to both specific and general sources of interference, showed significantly greater RI than did the OM pairs which were subject only to general List 1 response suppression.

The previous study (Birnbaum, 1970) contained four comparisons of recall of OM versus PR pairs,[3] and a total of five experimental groups was run in the present experiments. Considered as nine independent tests of the same hypothesis (Winer, 1962, pp. 43–44), the results demonstrate that recall of OM pairs is superior to recall of PR pairs, $\chi^2\ (18) = 46.58,\ p < .001$. If the overall retention losses in the experimental groups were attributable largely to suppression of the first-list response repertoire, as opposed to specific associative unlearning, there would have been little or no difference in the recall of OM and PR pairs. The findings thus warrant the conclusion that there was a substantial amount of specific associative unlearning. In addition, the fact that the omission of List 1 stimuli reduced but did not eliminate RI for the OM pairs (Exp. I) indicates that suppression of List 1 responses also made a significant contribution to first-list retention loss.

It is important to note that seven of the nine comparisons made above involved experimental groups that had learned two interpolated lists. It is possible that interpolation of a single A-C list would have produced less specific associative loss than was observed in the present experiments. It has been well established that multiple interpolation leads to greater RI than does single interpolation (Birnbaum, 1968; Howe, 1967; Postman, 1965). Whether the increases in RI with multiple interpolation resulted from increases in specific or general sources of interference, or both, remains an open question. The utilization of the OM versus PR design may provide an answer to this question and to similar questions concerning the locus of first-list retention losses under different conditions of transfer.

---

[3] 12S versus 10S, 12D versus 10D, 12S versus 8S, and 12D versus 8D.

## REFERENCES

Birnbaum, I. M. Unlearning as a function of second-list dominance. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 257–259.

Birnbaum, I. M. Response selection and retroactive inhibition. *Journal of Experimental Psychology*, 1970, **85**, 406–410.

Howe, T. S. Unlearning and competition in List 1 recall. *Journal of Experimental Psychology*, 1967, **75**, 559–565.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*, Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Postman, L. Unlearning under conditions of successive interpolation. *Journal of Experimental Psychology*, 1965, **70**, 237–245.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*. 1968, **7**, 672–694.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received September 4, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 193–197

# PREDICTION OUTCOME AND CHOICE REACTION TIME: STIMULUS VERSUS RESPONSE ANTICIPATION[1]

CHARLES P. WHITMAN AND E. SCOTT GELLER[2]

*Virginia Polytechnic Institute and State University*

Choice latency was significantly influenced by the outcome of *S*s' predictions on current and preceding trials in a discrete three-stimulus, two-response reaction time (RT) experiment. Three mutually exclusive categories of prediction outcome (PO) were correct stimulus and response prediction (CoS-CoR), incorrect stimulus and correct response prediction (InS-CoR), and incorrect stimulus and response prediction (InS-InR). Mean RTs of the current PO category, CoS-CoR, were significantly faster when the preceding PO was CoS-CoR. Mean RTs following an InS-CoR prediction were faster than RTs following an InS-InR prediction only when the preceding PO was InS-CoR. These results were interpreted as indications of the facilitation of the stimulus-identification or the response-selection components of the choice reaction process. The implications of an interaction between preceding PO and the complexity of the S–R relationship were discussed.

Beginning with a study by Bernstein and Reese (1965) and continuing through the present (e.g., Geller, Whitman, Wrenn, & Shipley, 1971), several choice reaction time (RT) investigators have required *S*s to predict the next stimulus following each choice reaction. The continued interest in this design has been generated by the consistent finding that *S*s react significantly faster when they correctly predict the stimulus than when their prediction was incorrect. Since *S*s' predictions are indications of expectation, many of the theories developed to explain the effect of making a correct prediction have used an expectancy hypothesis. One of the fundamental questions asked by expectancy theorists is whether *S*s' stimulus prediction results in preparation (or facilitation) to perceive a particular stimulus, to select a particular response, or to perform both of these cognitive components of the choice reaction process.

Hinrichs and Krainz (1970) concluded that shorter choice latencies following correct predictions are due to facilitating the stimulus identification component of the choice reaction process, while Keele (1969) found the prediction outcome effect to be attributed to facilitation of the response-selection process. There have also been studies which supported the view that only a combination of stimulus and response facilitation could account for the faster reactions to correctly predicted events (e.g., Bernstein, Shurman, & Forester, 1967). To resolve this issue, Hinrichs and Krainz (1970) used the three-stimulus, two-response design introduced by LaBerge and Tweedy (1964) to study the effects of prediction outcome on choice RT. Since two stimuli are mapped onto one response with this design, *S*s may correctly predict the response while incorrectly predicting the stimulus. With this paradigm, Hinrichs and Krainz (1970) found that choice RTs to correctly predicted stimuli were significantly faster than choice RTs to incorrectly predicted stimuli but that choice RTs to incorrectly predicted stimuli were not influenced by the outcome of the response prediction. In other words, when *S*s' predictions resulted in an incorrect anticipation of the stimulus but a correct anticipation of the response, their choice reactions were not reliably faster than those following an incorrect prediction of both

---

[1] Supported by the Small Grants Program of the College of Arts and Sciences, Virginia Polytechnic Institute and State University. This research was facilitated by college gift equipment contributed by Western Electric Company and the Chesapeake and Potomac Telephone Company. Special thanks are due to Richard F. Wrenn for constructing the logic equipment.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

stimulus and response. Therefore, the authors concluded that the facilitation of a choice reaction to a correctly predicted event was due to stimulus anticipation rather than to response preparation.

Recent studies by Whitman and Geller (1971a, 1971b) demonstrated that choice RT in a two-stimulus, two-response design is significantly influenced by the prediction outcome (PO) of trials preceding the current event as well as the PO of the current trial. The consistent finding of these studies was that reaction latencies to correctly predicted events were shorter if the preceding event was correctly predicted. However, when the current stimulus prediction was incorrect, the effect of the preceding PO on choice RT was dependent upon S–R compatibility (Whitman & Geller, 1971a). Specifically, with a compatible S–R relationship, the effect of the preceding PO on *S*s' latency to identify incorrectly predicted events was longer when the preceding prediction was correct than when it was incorrect. This finding was consistent with the continuous expectancy notions presented by Geller and Pitz (1970). On the other hand, with an incompatible S–R relationship, choice latency was shorter to those incorrectly predicted events preceded by correct POs than to those preceded by incorrect POs. Thus, the effects of preceding PO on choice RT to incorrectly predicted events were influenced by the level of compatibility or complexity of the S–R relationship.

The design of the present study, like that of Hinrichs and Krainz (1970), was a discrete choice reaction paradigm with three stimulus alternatives mapped onto two responses. Prior to each stimulus presentation, *S*s predicted which of the three stimuli they expected to occur. Since a correct response anticipation may not necessarily be also a correct stimulus anticipation with such a design, one purpose of the present study was to determine the relative influence of stimulus or response facilitation in preceding and current PO effects. The S–R compatibility condition used was quite different than those with which preceding PO effects were previously studied (Whitman & Geller, 1971a, 1971b). Although Hinrichs and Krainz assumed a three-stimulus, two-response design to be relatively compatible, the third stimulus would probably make such a design more complex (i.e., incompatible) than a two-stimulus design with natural S–R pairings (cf. Smith, 1968). Therefore, the three-stimulus design of the present study was more incompatible than the compatible condition of the Whitman and Geller (1971a) study but was perhaps more compatible than the two-stimulus condition of that study which employed unnatural S–R pairings. Consequently, because contrasting effects of preceding PO on choice RT to incorrectly predicted events were found for two conditions which represented opposite extremes of S–R compatibility (Whitman & Geller, 1971a), and since the S–R relationship of the present experiment was of moderate compatibility, the effect of preceding PO on choice RT following incorrectly predicted stimuli was of special interest in the present study.

## METHOD

*Subjects.*—Twenty *S*s (15 males, 5 females) participated in individual sessions. All *S*s were volunteers enrolled in introductory psychology classes at Virginia Polytechnic Institute and State University and received optional research credit for their participation.

*Apparatus and procedure.*—The three stimuli were the symbols ⊔, ⊓, and □ (verbally labeled "up," "down," and "square," respectively). These stimuli were presented by a $\frac{1}{2} \times 2$ in. digital readout mounted behind a one-way mirror. During each session, the room was dimly lit such that *S* could not see the readout unless it was illuminated.

The identification response consisted of pulling one of two microswitches, one being mounted in a handle for *S*s' left hand and the other in a right-hand trigger. The up and down stimuli each occurred on 25% of the 200 trials and were paired with the same response. Since the square stimulus occurred on the remaining 50% of the trials and was identified by the second response, the two responses were equally probable. Ten *S*s pulled the right trigger to identify ⊔ or ⊓ and the left trigger to identify □. The reverse S–R relationship (i.e., a left response for ⊔ or ⊓ and a right response for □) was used by the remaining 10 *S*s.

The task instructions urged *S*s to concentrate while making their predictions and emphasized the importance of both speed and accuracy when identifying each stimulus. The *S*s were told that the stimuli were preprogrammed on punched paper tape but were not informed of the relative frequency of each stimulus. At the beginning of the session, each *S* was given 10 practice trials during which questions concerning his task were answered.

Following *S*s' verbal prediction, a brief warning buzzer sounded which signaled the beginning of a random time interval of less than 2 sec. to precede the stimulus presentation. The *S*s' choice reaction turned off the stimulus and stopped a Hunter Klockcounter. The predictions, the errors, and the response latencies were manually recorded. Errors (i.e., anticipatory responses or incorrect identifications) occurred on less than 5% of the trials and were not considered in the analysis.

## RESULTS

Considering the anticipations of stimuli and responses separately in the three-stimulus paradigm resulted in 3 types of a prediction outcome (PO). A correct stimulus prediction (CoS) necessarily implied a correct response prediction (CoR), whereas an incorrect stimulus prediction (InS) was either a correct response prediction (CoR) or an incorrect response prediction (InR). Consequently, the 3 PO categories and their abbreviations were as follows: correct prediction of stimulus and response (CoS-CoR), incorrect stimulus and correct response prediction (InS-CoR), and incorrect prediction of both stimulus and response (InS-InR). These 3 mutually exclusive PO categories defined 3 categories of current PO and, within each current PO condition, 3 subcategories of a preceding PO. Thus, the choice RT of Trials 2 through 200 was entered in one of nine categories determined by the 3 × 3 factorial of preceding and current PO. For example, the entry of a given *S* in the category representing current POs of CoS-CoR and preceding POs of InS-CoR was the mean latency over all trials on which that *S* made a correct stimulus and correct response prediction and on the immediately preceding trial had made an incorrect stimulus and correct response prediction. No *S* had fewer than six latency entries to be averaged in any PO category.

The mean choice RT for each PO category is depicted in Fig. 1. Given a current PO of correct stimulus and correct response (CoS-CoR), choice reactions were fastest when the preceding PO was CoS-CoR but were relatively uninfluenced by a preceding PO of incorrect stimulus and correct response (InS-CoR). In contrast, given a current PO of InS-CoR, choice latency was shortest when the preceding PO was InS-CoR but was relatively uninfluenced by a preceding PO of CoS-CoR. Finally, when the current PO was incorrect for both stimulus and response (InS-InR), choice RT was apparently not affected by preceding PO. The overall analysis of variance, a 3 (Current PO) × 3 (Preceding PO), indicated significant main effects of both current PO, $F\ (2,\ 38) = 40.00$, and preceding PO, $F\ (2,\ 38) = 9.80$, and a significant interaction between current PO and preceding PO, $F\ (4,\ 76) = 6.69$, all $ps < .001$. The interaction may be largely accounted for by the finding that the influence of preceding PO was dependent upon the current PO. In order to examine

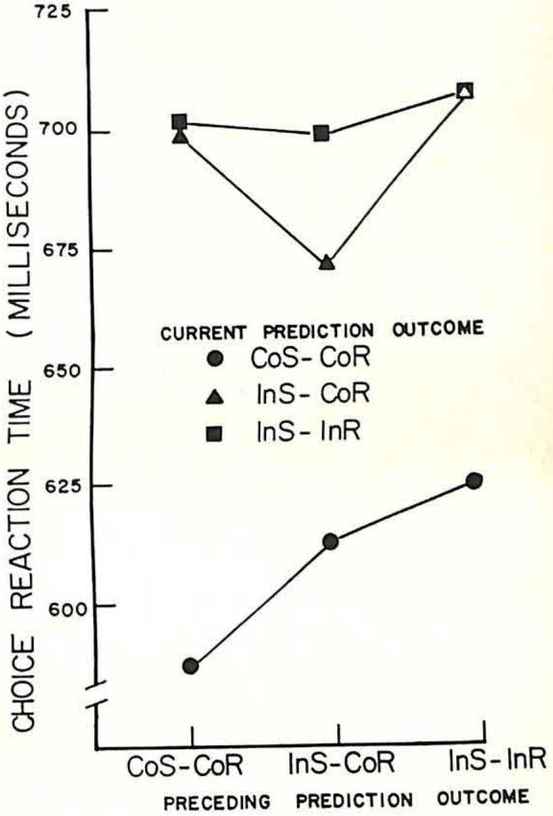


FIG. 1. Choice reaction time as a function of preceding and current prediction outcome categories. (CoS-CoR refers to a correct stimulus and response prediction, InS-CoR an incorrect stimulus and correct response prediction, and InS-InR an incorrect prediction of both stimulus and response.)

TABLE 1

ABSOLUTE $t$ VALUES BETWEEN THE PRECEDING PREDICTION OUTCOME CATEGORIES WITHIN EACH CURRENT PREDICTION OUTCOME CATEGORY

| Current prediction outcome | Preceding prediction outcome comparisons | | |
|---|---|---|---|
| | CoS-CoR vs. InS-CoR | CoS-CoR vs. InS-InR | InS-CoR vs. InS-InR |
| CoS-CoR | 3.05* | 4.75* | 1.45 |
| InS-CoR | 3.11* | .96 | 3.22* |
| InS-InR | .61 | 1.03 | 1.70 |

Note.—Degrees of freedom = 19.
* $p < .01$ (two tailed).

this complex interaction, the differences between latencies of individual PO categories were examined by a series of post hoc (i.e., two-tailed) $t$ tests.

As demonstrated in Table 1, when the current PO was CoS-CoR, a correct prediction on a preceding trial resulted in a faster choice reaction only when both the stimulus and the response had been correctly predicted (CoS-CoR). An incorrect anticipation of the stimulus paired with a correct anticipation of the response (InS-CoR) on a preceding trial did not influence choice RT to a correctly predicted stimulus (CoS-CoR). Such is indicated by the reliable difference ($p < .01$) between mean latencies for preceding PO categories CoS-CoR and InS-CoR and by the lack of a significant difference between the mean choice RTs of preceding PO categories InS-CoR and InS-InR when the current PO was CoS-CoR. On the other hand, given a current PO of InS-CoR, a correct anticipation of only the response (InS-CoR) on a preceding trial significantly shortened *S*s' choice RT on the subsequent trial (i.e., the differences between mean latencies for preceding PO category InS-CoR and preceding PO categories CoS-CoR and InS-InR were significant, $p$s $< .01$). The results of the $t$ tests also showed that choice RT following an incorrect stimulus and incorrect response prediction (InS-InR) was not significantly influenced by the PO of the preceding trial.

It is of considerable interest to examine the data presented in Fig. 1 with a slightly different approach. That is, instead of studying the effects of preceding PO within the three categories of current PO, it is worthwhile to consider the effects of current PO within the three categories of preceding PO. In Fig. 1, this corresponds to selecting a preceding PO category on the abscissa and comparing the effects of current PO category on the ordinate. Given a preceding PO of CoS-CoR or of InS-InR, there was no significant difference between the choice RTs of the current PO categories: InS-CoR and InS-InR (both $t$s $> .3$). However, when only the response was correctly predicted on the preceding trial (InS-CoR), there was a reliable difference between current PO categories InS-CoR and InS-InR, $t = 2.18$, $p < .05$. If the preceding PO was CoS-CoR or InS-InR, then correctly predicting only the response (InS-CoR) on the current trial did not influence choice RT.

## DISCUSSION

Considering the effects of current PO within each of the three categories of preceding PO, there was a significant facilitating effect of correctly predicting both the stimulus and response in every case. In addition, given a preceding PO of correct stimulus and response or of incorrect stimulus and response, there were no significant differences between the current PO categories, incorrect stimulus-correct response and incorrect stimulus and response. This observation of no effect of correctly predicting only the response on a current trial is consistent with the findings of Hinrichs and Krainz (1970). However, when the preceding PO was incorrect stimulus-correct response, there was a reliable effect of currently predicting only the response correctly. The implications of these results are in contrast with the conclusions of Hinrichs and Krainz (1970), that PO effects are due entirely to stimulus anticipation. However, since collapsing across the preceding PO categories would result in mean latencies consistent with the findings of Hinrichs and Krainz, the inference to be made is that the PO of the preceding trial influences the extent to which response anticipation shortens choice latencies to correctly predicted events. Clearly though, stimulus rather than response anticipation was demonstrated to be the primary choice RT determinant in two of the three preceding prediction outcome categories.

Considering the effects of preceding PO within each of the current PO categories, a noteworthy result was the one which indicated that stimulus anticipation may be the more prominent factor contributing to the effect of preceding PO on choice RT. That is, the comparison between the latencies of the PO categories preceding a current PO of correct stimulus and response indicated that choice RT was greatly facilitated by a preceding PO of correct stimulus and response. However, there was no significant difference between mean latencies of the preceding PO categories incorrect stimulus-correct response and incorrect stimulus-incorrect response. Thus, when the current PO was a correct stimulus and response, the effect of preceding PO was apparently due to stimulus rather than response anticipation. In contrast, when the current PO was incorrect stimulus-correct response, the observed effect of preceding PO did implicate response preparation. Specifically, given a current PO of incorrect stimulus-correct response, only the preceding PO of incorrect stimulus-correct response facilitated choice reactions. Although the theoretical implications of this finding are not clear, it is important to note that correct response preparation did significantly influence choice RT for this particular preceding PO category.

It should be noticed that choice RT to an incorrectly predicted stimulus and response was not significantly influenced by preceding PO. Such a result is unlike any previous observations of preceding PO effects on choice RT to incorrectly predicted events (Whitman & Geller, 1971a, 1971b). Perhaps this unusual result may be explained by the differences in the compatibility of the S–R relationships used. That is, the three-stimulus, two-response design of the present study is less compatible than the compatible condition of the Whitman and Geller (1971a) study but is perhaps more compatible than the incompatible condition of that study (cf., Smith, 1968). Then, combining the results of Whitman and Geller (1971a) with those of the present study: correct POs preceding an incorrect PO inhibit choice RT when the S–R relationship is extremely compatible, have no effect on choice RT for a moderately compatible condition, and facilitate choice RT if the S–R relationship is extremely incompatible. The suggestion is that the effect of preceding PO on choice reaction to incorrectly predicted events ranges from inhibition to facilitation as the S–R relationship becomes more incompatible.

In summary, the present study found choice RT in a three-stimulus, two-response design to be significantly influenced by the outcome of *Ss'* prediction on the trial preceding the identification response and that the facilitation due to a correct prediction may be attributed to either stimulus or response expectancy. Although stimulus effects were of a more striking nature, response effects were evident in the effect of current PO when the preceding PO involved incorrectly predicting the stimuli while correctly predicting the response. Thus, the results of the present study indicated the necessity of considering preceding as well as current PO in research attempting to discern between stimulus and response influences in PO effects. The effect of preceding PO on choice latency to correctly predicted stimuli was consistent with the results of previous studies. However, the absence of a significant effect of preceding PO on choice RT to incorrectly predicted events was inconsistent with earlier findings and demonstrated the sensitivity of the preceding PO effect to S–R compatibility factors.

## REFERENCES

Bernstein, I. H., & Reese, C. Behavioral hypotheses and choice reaction time. *Psychonomic Science*, 1965, **3**, 259–260.

Bernstein, I. H., Schurman, D. L., & Forester, G. Choice reaction as a function of stimulus uncertainty, response uncertainty, and behavioral hypotheses. *Journal of Experimental Psychology*, 1967, **74**, 517–524.

Geller, E. S., & Pitz, G. F. Effects of prediction, probability, and run length on discrete choice reaction speed. *Journal of Experimental Psychology*, 1970, **84**, 361–367.

Geller, E. S., Whitman, C. P., Wrenn, R. F. & Shipley, W. G. Expectancy and discrete reaction time in a probability reversal design. *Journal of Experimental Psychology*, 1971, **90**, 113–119.

Hinrichs, J. V., & Krainz, P. L. Expectancy in choice response time: Anticipation of stimulus or response? *Journal of Experimental Psychology*, 1970, **85**, 330–334.

Keele, S. W. Repetition effect: A memory dependent process. *Journal of Experimental Psychology*, 1969, **80**, 243–248.

LaBerge, D., & Tweedy, J. R. Presentation probability and choice time. *Journal of Experimental Psychology*, 1964, **68**, 477–481.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

Whitman, C. P., & Geller, E. S. Prediction outcome, S-R compatibility, and choice reaction time. *Journal of Experimental Psychology*, 1971, **91**, 299–304. (a)

Whitman, C. P., & Geller, E. S. Runs of correct and incorrect predictions as determinants of choice reaction time. *Psychonomic Science*, 1971, **23**, 421–423. (b)

(Received September 7, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 198–204

# INTERHEMISPHERIC EFFECTS ON CHOICE REACTION TIMES TO ONE-, TWO-, AND THREE-LETTER DISPLAYS[1]

CARLO UMILTÁ,[2] NANCY FROST,[3] AND RAY HYMAN[4]

*University of Oregon*

In Exp. I, 16 normal right-handed *S*s saw letters presented either to the right or left of a central fixation mark. The *S*s pressed a key when those stimuli designated as positive appeared, and did not respond to all others. Both right- and left-hand responding were used. Analysis of correct responses indicated a nonsignificant left hemifield superiority with one letter, and significantly faster responding to right hemifield presentations for two- and three-letter displays. In Exp. II, 10 *S*s responded to single letters, differing simultaneously in visual and phonemic discrimination. Reaction times can be interpreted to indicate faster phonemic discriminations of vowel and stop consonant sounds for left and right hemifields, respectively. Results are discussed in terms of hemispheric specialization and interhemispheric transmission of visual information.

Recently, many studies have dealt with the subject of cerebral dominance in man by using the choice reaction time (RT) paradigm. Essentially, latencies are measured for discriminative responses to a visual pattern whose recognition is thought to be subserved predominantly by one particular hemisphere. In such tasks, stimuli presented in the visual hemifield projecting to the specialized hemisphere should be responded to faster than stimuli presented in the other visual hemifield, since in the first case the visual input is directly received and processed by the specialized hemisphere. Motor responses are often preferred to vocal responses because the latter are probably initiated only by the left hemisphere (Gazzaniga & Hillyard, 1971; Sperry, 1968). This implies that an RT advantage should occur for stimuli presented in the right visual hemifield due merely to the required verbal response.

The results so far achieved by using motor responses indicate a left hemifield (i.e., right hemisphere) superiority for responses to pictures of human faces (Geffen, Bradshaw, & Wallace, 1971; Rizzolatti, Umiltá, & Berlucchi, 1971). Such data support the concept that the analysis, storage, and retrieval of visual material are better performed by the right hemisphere when stimuli are coded in terms of certain spatial and configurational attributes. Likewise, one would expect a right hemifield (i.e., left hemisphere) superiority for responses to verbal stimuli, since the analysis, storage, and retrieval of visual material endowed with verbal meaning is most efficiently carried out by the left hemisphere (see review by Geschwind, 1970). However, the experiments concerning this subject have yielded some contradictory findings. Rizzolatti et al. (1971) and Geffen et al. (1971) found an advantage for right hemifield presentation by using single letter and number stimuli, respectively. To the contrary, a left hemifield superiority for responses to single letters has been reported by Klatzky (1970), Klatzky and Atkinson (1971), and Moskovich.[5] These authors did, however, find a right hemifield superiority for responses to single letters when the memory load was increased from one letter to six (See Footnote 5), or when *S*s were forced to retrieve the first letter of a word in order to perform the correct response (Klatzky, 1970; Klatzky & Atkinson, 1971).

---

[1] Funds for *S*s, computer time, and data analysis were provided by the Advanced Research Projects Agency of the Department of Defense, monitored by the Air Force Office of Scientific Research under Contract No. F44620-67-0099 to the University of Oregon. The authors thank Martha Snyder for her assistance in running *S*s and analyzing data.

[2] Now at the Institute of Psychology of the University of Bologna, Bologna, Italy.

[3] Now at the University of Washington, Seattle, Washington.

[4] Requests for reprints should be sent to Ray Hyman, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

[5] Moscovitch, M. A. A choice reaction time study measuring the verbal abilities of the minor hemisphere. Manuscript in preparation, 1972.

The work of Posner (see review, 1969) suggests that these discrepancies might be accounted for, since it is known that *S*s can perform the task by matching single letters on either a physical or a name basis. According to this concept, matches involving physical codes show an advantage for stimuli presented to the left hemifield (i.e., directly to the right hemisphere), while matches involving a name code show an advantage for stimuli presented to the right hemifield (i.e., directly to the left hemisphere).

Up to this time, two major hypotheses have been proposed to explain, in terms of cerebral specialization, the observed differences between the two hemifields with regard to the speed of response to various stimuli (Filbey & Gazzaniga, 1969; Gazzaniga & Hillyard, 1971; Geffen et al. 1971). The first suggests that the superiority of one visual hemifield for some visual stimuli is due to the fact that stimuli thus projected reach directly the specialized hemisphere where they are immediately processed; while, if they reach the brain from the other visual hemifield, they have to cross from the nonspecialized hemisphere to the other side of the brain for analysis. Therefore, according to this hypothesis, the difference in response time to stimuli presented in the two visual hemifields reflects the time necessary for interhemispheric transmission of visual information from the nonspecialized hemisphere to the specialized one.

The second hypothesis suggests that the nonspecialized hemisphere can also analyze visual material presented to the corresponding visual hemifield, but takes longer than the specialized hemisphere to do so. Therefore, according to this hypothesis, the difference in response time between the two visual hemifields merely reflects the difference in processing speed between the two hemispheres.

The present experiments are designed (*a*) to investigate the discrepancies in the results obtained with verbal material, (*b*) to confront the two explanatory hypotheses which have so far been reported, and (*c*) to discover whether the difference in RT between the two visual hemifields increases with the complexity of the required discrimination.

## EXPERIMENT I

### *Method*

*Subjects.*—Ten male and 6 female *S*s served in Exp. I. All were right-handed and had normal or corrected vision; all were uninformed as to the purpose of the experiment and were paid $1.50 an hour for their participation. Each *S* was run individually in five 80-min. sessions, on 5 consecutive days.

*Apparatus.*—Stimulus presentation, response scoring, and all interval and latency timing were controlled by a DEC PDP-9 computer. Twenty capital English letters comprised the stimulus set. These letters were displayed 1, 2, or 3 at a time on a 17.5 × 23 cm. oscilloscope screen. Directly beneath the center of the screen was a single key which *S* could press with either his right or his left index finger. Directly in front of the oscilloscope was a chin rest which positioned *S*'s eyes at the same height as, and approximately 34.3 cm. away from, the center of the screen. In the middle of the screen, a fixation cross (6 × 6 mm.) appeared 1 sec. before stimulus onset and remained throughout a 100-msec. stimulus display (200-msec. during the first, practice, session) and a 100-msec. masking interval (during which a solid grid of light, 6 × 8 mm., erased each stimulus letter; the masking interval insured that the oscilloscope's decay time would not lengthen the actual stimulus display). Each stimulus letter appeared 6 × 8 mm. in size, with 4-mm. spacing between those stimuli presented in rows of 2 or 3 letters. Between the center of the fixation cross and the center of the 1- and 3-letter displays there was a distance of 48 mm. (i.e., a visual angle of approximately 8°2′); 43 mm. (i.e., approximately 7°12′) separated the center of the fixation cross and the center of the 2-letter displays.

*Design.*—The task was a "go-no go" discrimination: *S* was required to press the key as soon as possible following the appearance of certain (positive) stimuli and not to press the key following the appearance of other (negative) stimuli. A new trial began 3 sec. after the end of the last, whether or not *S* emitted a response.

Three different experimental conditions were used. Each condition consisted of eight stimuli, randomly occurring with unequal probability. Two positive stimuli appeared with .26 probability; while the probability of each of the six negative stimuli was .08. In the one-letter condition (CI) the following letters appeared one at a time: C, T, B, F, D, E, M, and O. For eight *S*s, the positive stimuli were C and

TABLE 1

EXPERIMENT I: OVERALL MEAN REACTION TIME, STANDARD DEVIATION, AND NUMBER OF ERRORS

| Cond. | Statistic | Side of stimulus | | Interaction: Side of stimulus and responding hand | | | |
|---|---|---|---|---|---|---|---|
| | | RF | LF | RF-RH | RF-LH | LF-RH | LF-LH |
| C1: one-letter display | $\bar{X}$ | 479.0 | 469.2 | 481.9 | 476.0 | 474.7 | 463.6 |
| | *SD* | 108.4 | 103.8 | 112.3 | 101.7 | 105.6 | 99.7 |
| | Errors | 76 | 58 | 43 | 33 | 24 | 34 |
| C2: two-letter display | $\bar{X}$ | 604.0 | 649.1 | 599.0 | 609.0 | 646.5 | 651.7 |
| | *SD* | 107.2 | 139.6 | 111.5 | 101.3 | 139.3 | 139.9 |
| | Errors | 281 | 472 | 123 | 158 | 248 | 224 |
| C3: three-letter display | $\bar{X}$ | 622.9 | 685.2 | 618.4 | 627.4 | 680.8 | 689.6 |
| | *SD* | 150.9 | 169.9 | 144.1 | 152.2 | 166.9 | 172.0 |
| | Errors | 396 | 520 | 201 | 195 | 241 | 279 |

Note.—Reaction times in milliseconds. RF, right visual hemifield; LF, left visual hemifield; RH, right hand; LF, left hand.

T, and for the other eight *S*s, B and F were positive. In the two-letter condition (C2) the following letter pairs appeared one at a time: AW, NR, AG, AR, AH, NG, NH, and NW. The positive stimuli were always AW and NR. In the three-letter condition (C3) the following letter triplets were presented one at a time: ISU, ZPL, IPL, IPU, ISL, ZPU, ZSL, and ZSU. ISU and ZPL were always positive. The 20 letters were randomly assigned to conditions, pairings within C2 and C3, and to the positive and negative stimulus sets. The organization of the letter pairs and triplets in C2 and C3 forced *S* to read all letters in order to correctly respond to the positive stimuli.

By following *E*'s instructions, *S* responded to a series of trials with either his right or his left index finger. Thus four Hand × Visual Hemifield combinations were possible in every condition: right hemifield stimulus and right-hand response (RF-RH); right hemifield stimulus and left-hand response (RF-LH); left hemifield stimulus, right-hand response (LF-RH); and left hemifield stimulus, left-hand response (LF-LH).

*Procedure.*—During testing, *S* sat alone in a darkened room, separated from *E* and the computer. The 80-min. session was divided into 12 blocks of 50 trials each. After every 10 trials *S* has a rest of approximately 15 sec., during which the number of correct discriminations made on the last 10 trials was displayed in the center of the screen. On Days 1 and 2 (practice sessions, whose data have not been used in the analyses), Cond. C1, C2, and C3 were run in respective order. On Days 3, 4, and 5 the order of C1, C2, and C3 were varied for each *S* according to a Latin square design. Within each condition, there were four blocks of trials (one for each of the Hand × Visual Hemifield combinations) whose order of presentation was varied randomly. Trials were blocked both by response hand and by visual hemifield in order to minimize effects of incompatibility in RF-LH and LF-RH combinations. At the end of every block, *S* had a 1-min. rest during which he could read information pertinent to the next block of trials; the screen displayed the positive stimuli and the side of the fixation cross on which all stimuli would appear. The *E* then instructed *S* to respond with either right or left hand during the block of trials to follow. Within each block of trials, the stimuli were presented, according to the above mentioned probabilities, in random sequences monitored by the computer.

At the beginning of every session the instructions stressed that it was of paramount importance for *S* to keep a good fixation during the task.

### *Results*

An RT[6] and an error analysis of variance, using a three-factor repeated-measurement design, were performed for each of the three conditions. In each analysis, the factors were: hands (H) with two levels, right and left; fields (F), right or left visual hemifield; and days (D) with levels for the third, fourth, and fifth experimental sessions. The results are summarized in Table 1.

---

[6] Significant correlations between mean RT for both hemifield presentations ($X$) and the difference in RT for the two hemifields ($Y$) indicated that a log transformation was appropriate in all further analysis of RT data. Therefore, analyses of variance for RTs were performed on both raw and transformed times. No essential differences resulted in the two methods of analysis, except for slightly enhanced significance using the transformed scores. Only those *F*s for transformed RTs are reported.

There was a 9.8-msec. advantage for the left visual hemifield, $F$ (1, 15) = .794, $p > .20$, and errors were fewer for left hemifield presentation, $F$ (1, 15) = 1.869, $p > .10$. No source of variance in either RT or error analyses attained significance; half of the *S*s in the Cl condition showed faster RTs in the left hemifield and half of the *S*s were faster to right hemifield presentation.

In Cond. C2, Source F is significant for both RT and error data. There was a 45.1-msec. advantage for the right visual hemifield, $F$ (1, 15) = 22.585, $p < .001$, and errors were fewer for right hemifield presentation, $F$ (1, 15) = 21.342, $p < .001$. All 16 *S*s were faster with right hemifield stimulus presentation. The H × F interaction attained significance in the error analysis; errors were fewer when the response was performed by the ipsilateral hand, $F$ (1, 15) = 5.440, $p < .05$.

In Cond. C3, Source F is again significant for both RT and error analyses. There was a 62.3-msec. advantage for the right visual hemifield, $F$ (1, 15) = 10.841, $p < .005$, and errors were fewer for right hemifield presentation, $F$ (1, 15) = 7.253, $p < .025$. The RTs for 14 of the 16 *S*s were consistent with these results. No other sources of variance for either C2 or C3 analyses attained significance.

An analysis of variance for a one-factor, repeated-measurement design was performed to test the difference between the right hemifield advantage noted for C2 and C3. While the left–right hemifield difference was greater in C3 than in C2 for 10 of the 16 *S*s, the right hemifield advantage was not significantly larger for C3, $F$ (1, 15) = 2.018, $p > .10$.

## EXPERIMENT II

The one-letter condition has failed to support either those results which reported an advantage for the right visual hemifield or those which demonstrate an advantage for the left visual hemifield. The lack of significant and consistent results in Cond. C1 may be explained by the hypothesis which states that *S*s can process single letter stimuli on either the name or shape level. However, such an explanation is unsatisfactory without a better understanding of those conditions under which name matches or physical matches take place. Another hypothesis is suggested by Crowder (1971): that the left (or dominant) hemisphere can process better the phonemic information carried by stop consonants; whereas, vowel sounds are better processed by the right (or nondominant) hemisphere. If this holds also for visually presented material, the implication for the present paradigm is that discriminations involving letters differentiated, when pronounced, only by consonant sounds (e.g., "dee," "tee," and "bee") should be performed better by the left cerebral hemisphere. Discriminations involving letters which differ, when pronounced, in terminal vowel sounds should be performed better by the right cerebral hemisphere. Cond. C1, it should be noted, is mixed from this point of view (e.g., C and O can be discriminated by terminal vowels, while C and D cannot).

The second experiment was designed to investigate both the Posner (1969) and the Crowder (1971) hypotheses. Two conditions were used, one in which all stimuli contained the same ("ee") vowel sound but in which positive and negative stimuli were easily discriminable by shape. The second condition contained stimuli which were highly similar in shape but which were distinguishable, positive from negative, by both consonant and vowel sounds. It was reasoned that if name–shape matches were operating in the one-letter condition, then the former (visually distinguishable) stimuli should produce a right cerebral hemisphere advantage; and vice versa for visually confusable stimuli whose names are easily discriminated. However, if the advantage of one hemisphere over another was determined at a phonemic level, then the former stimulus condition should produce the reverse effect (i.e., a left hemisphere advantage).

### *Method*

*Subjects.*—Ten new right-handed and normally sighted *S*s, 6 females and 4 males, served in Exp. II. All were paid $1.50 an hour for their participation,

TABLE 2

EXPERIMENT II: OVERALL MEAN REACTION TIME, STANDARD DEVIATION, AND NUMBER OF ERRORS

| Cond. | Statistic | Side of stimulus | | Interaction | | | |
|---|---|---|---|---|---|---|---|
| | | | | Side of stimulus and responding hand | | | |
| | | RF | LF | RF-RH | RF-LH | LF-RH | LF-LH |
| C4: shape-consonant discrimination | $\bar{X}$ | 508.5 | 533.6 | 503.4 | 513.6 | 532.6 | 534.7 |
| | *SD* | 195.0 | 150.6 | 198.8 | 176.0 | 140.8 | 155.8 |
| | Errors | 55 | 65 | 27 | 28 | 35 | 30 |
| C5: name-vowel discrimination | $\bar{X}$ | 507.0 | 487.6 | 516.2 | 497.8 | 489.1 | 486.0 |
| | *SD* | 127.4 | 98.7 | 132.3 | 118.4 | 99.0 | 98.2 |
| | Errors | 291 | 212 | 129 | 162 | 113 | 99 |

Note.—Reaction times in milliseconds. RF, right visual hemifield; LF, left visual hemifield; RH, right hand; LH, left hand.

and none were informed of the purpose of the experiment. Each *S* was run individually for 4 consecutive days in sessions of 60 min. each.

*Apparatus.*—The apparatus described for Exp. I was also used in the present experiment. However, two differences must be noted: (*a*) six capital letters comprised the stimulus set for Exp. II; and (*b*) a 200-msec. stimulus display time was used only during the first block of the first and third sessions.

*Design.*—Two experimental conditions were used. In each condition, the probability of the two positive stimuli was .26; while the probability of the two negative stimuli was .24. In the shape-consonant discrimination (C4) condition, four letters appeared one at a time: D, G, T, and Z. The positive stimuli were D and G. In Cond. C5, name-vowel discrimination, four letters appeared one at a time: D, G, O, and Q. The positive stimuli were again D and G. The same four Hand × Visual Hemifield combinations were used: RF-RH, RF-LH, LF-RH, and LF-LH.

*Procedure.*—Eight blocks of 50 trials were run per day, plus two practice blocks (whose data have not been used in the analysis) at the beginning of Days 1 and 3. Five *S*s ran in Cond. C4 on Days 1 and 2 and in Cond. C5 on Days 3 and 4. For the other five *S*s, the order was reversed. Within each condition, there were two blocks of trials for every Hand × Visual Hemifield combination, the order of which was varied randomly. In all other respects, the procedure for Exp. II replicated that of the first experiment.

## *Results*

The RT data (see Table 2) were tested by an analysis of variance for mixed designs. The four within-*S* factors were: hand (H), either right or left; field (F), either right or left hemifield presentation; Cond. (C), C4 or C5; and days (D), first or second, within each condition run. Factor O (Order, i.e., whether C4 preceded or followed C5) was a between-*S* variable. Owing to a technical fault the error data were lost for two *S*s, and thus no formal error analysis was made. However, the remaining error data are also presented in Table 2.

The F × C interaction proved to be significant, $F\ (1,\ 8) = 9.653,\ p < .025$. The right visual hemifield had a 25.1-msec. advantage over the left in the C4 condition (8 of the 10 *S*s demonstrated this advantage); and the left visual hemifield was 19.4 msec. faster than the right in the C5 condition (7 of the 10 *S*s were consistent with this result). The error data show the same interaction: errors are fewer in the right hemifield for C4 and in the left hemifield for C5. Minor information is supplied by other significant sources of variance. The factor D, $F\ (1, 8) = 24.812, p < .005$, shows that RTs were 26 msec. faster on the second day—an obvious practice effect. The D × O interaction, $F\ (1, 8) = 9.516$, $p < .025$, indicates that the practice effect was larger for those *S*s who ran in Cond. C4 first (42 msec. vs. 10 msec.). On the other hand, it is hard to interpret significant O × C, $F\ (1,\ 8) = 7.890,\ p < .025$, and O × H, $F\ \ (1,\ \ 8) = 19.185,\ \ p < .005$. According to O × C, RTs are faster for Cond. C4 than for C5 only when the former was run first. According to O × H, the left-hand responding was faster only when C4 was run first. No other source of variance was significant.

## DISCUSSION

The results of the second experiment indicate that Crowder's (1971) hypothesis, that vowel and consonant discrimination are hemispherically specialized tasks, is one source of the discrepancies found in interhemispheric studies using single-letter stimuli and motor responding. When the task involves discriminations between letters differing only in stop consonant sounds, the RT is faster when information is sent directly to the left hemisphere; but when the task involves discriminations between letters differing in vowel sounds, the RT data favor the right cerebral hemisphere. Nevertheless, it is quite likely that other sources of discrepancies operate in other experimental conditions (e.g., physical vs. name matches, according to Posner's (1969) terminology). White (1971), for example, suggests that certain visual discriminations (verticals) are easier than others and hence make some letters (e.g., T) more amenable to right hemisphere processing.

On the other hand, the right hemifield is clearly superior when *S*s are forced to read two or three letters in order to perform correct discriminations, suggesting a left hemisphere specialization. This finding is also compatible with a theory of acquired reading habits which operate according to a postexposural trace-scanning mechanism (Heron, 1957; White, 1969). However, recent evidence, reported elsewhere (McKeever & Huling, 1970a; 1970b) tends to discredit such a theory. Furthermore, a right hemifield superiority for motor responses to letters is also compatible with a motor theory of recognition. Klapp (1971) has shown that in recognition tasks the response latency depends on number of syllables which are needed to pronounce numerical or word stimuli; his finding holds for both motor and vocal responding. Thus, if recognition of verbal material needs some sort of implicit speech, a right hemifield superiority might actually reflect the left hemisphere specialization for speech production. At present, there are no decisive data to support the hemispheric specialization hypothesis against the other two; but neither the reading habit hypothesis nor the implicit speech hypothesis can explain the results of Exp. II. Therefore, we currently prefer the hemispheric specialization notion which is consistent with both Exp. I and Exp. II results.

Assuming that our results are due to hemispheric specializations, it is still unclear whether RT advantages reflect interhemispheric transmission time or differences in processing speed between specialized and nonspecialized hemispheres. Because the control of finger movements is almost completely lateralized in the opposite hemisphere, the first hypothesis suggests one definite ordering of response speed among the four Hand × Visual Hemifield combinations. One combination (RF-RH or LF-LH, depending upon the side of the specialized hemisphere) implies no interhemispheric crossing in order to perform the motor reaction to the stimulus. Another combination (again RF-RH or LF-LH, depending upon the side of the nonspecialized hemisphere) implies two interhemispheric crossings. The remaining two combinations, LF-RH and RF-LH, always imply one crossing. The Hand × Field combinations for all the data reported here, including those for Cl where no hemispheric preference is indicated, are consistent with the transmission hypothesis (see Tables 1 and 2). The fastest RTs have been found for that combination which implies no interhemispheric crossing; while the slowest RTs have been found for that combination which implies two interhemispheric crossings. The RTs for those combinations implying one crossing are intermediate, but always faster when the interhemispheric crossing would initiate a motor response than when the crossing would occur prior to stimulus processing. These data, while not conclusive, do support a theory which states that increases in RT are due to interhemispheric transmission of visual information.

Since no significant difference was found between RT advantage in two- and three-letter conditions, little can be concluded about the relationship between time and complexity of visual discriminations. The data do suggest, however, the possible existence of such a relationship. The difference in RT between the two hemifield conditions tends to increase with the complexity of the required discrimination: 9.8, 19.4, and 25.1 msec. for one-letter; 45.1 msec. for two-; and 62.3 msec. for three-letter stimuli.

## REFERENCES

CROWDER, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 587–596.

FILBEY, R. A., & GAZZANIGA, M. S. Splitting the normal brain with reaction time. *Psychonomic Science*, 1969, **17**, 335–336.

GAZZANIGA, M. S., & HILLYARD, S. A. Language and speech capacity of the right hemisphere. *Neuropsychologia*, 1971, **9**, 281–288.

GEFFEN, G., BRADSHAW, J. L., & WALLACE, G. Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. *Journal of Experimental Psychology*, 1971, **87**, 415–422.

GESCHWIND, N. The organization of language and the brain. *Science*, 1970, **170**, 940–944.

HERON, W. Perception as a function of retinal locus and attention. *American Journal of Psychology*, 1957, **70**, 38–48.

KLATZKY, R. L. Interhemispheric transfer of test stimulus representation in memory scanning. *Psychonomic Science*, 1970, **21**, 201–203.

KLATZKY, R. L., & ATKINSON, R. C. Specialization of the cerebral hemispheres in scanning for information in short-term memory. *Perception & Psychophysics*, 1971, **10**, 335–338.

KLAPP, S. T. Implicit speech in number recognition. Paper presented at the meetings of the Western Psychological Association, San Francisco, California, April 1971.

MCKEEVER, W. F., & HULING, M. D. Lateral dominance in tachistoscopic word recognitions of children at two levels of ability. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 600–604. (a)

MCKEEVER, W. F., & HULING, M. D. Left-cerebral hemisphere superiority in tachistoscopic word-recognition performances. *Perceptual & Motor Skills*, 1970, **30**, 763–766. (b).

POSNER, M. I. Abstraction and the process of recognition. In J. T. Spence & G. Bower (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969.

RIZZOLATTI, G., UMILTÁ, C., & BERLUCCHI, G. Opposite superiorities of the right and left cerebral hemispheres in discriminative reaction time to physiognomical and alphabetical material. *Brain*, 1971, **94**, 431–442.

SPERRY, R. W. Hemispheric deconnection and unity in conscious awareness. *American Psychologist*, 1968, **23**, 723–733.

WHITE, M. J. Laterality differences in perception: A review. *Psychological Bulletin*, 1969, **72**, 387–405.

White, M. J. Visual hemifield differences in the perception of letters and contour orientation. *Canadian Journal of Psychology*, 1971, **25**, 207–212.

(Received September 15, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 205–212

# INTERPOLATED ACTIVITY AND RESPONSE MECHANISMS IN MOTOR SHORT-TERM MEMORY[1]

DON TRUMBO,[2] FRANCIS MILONE,[3] AND MERRILL NOBLE

*Pennsylvania State University*

Eighty *S*s in four groups attempted to reproduce criterion movements in which trials varied in criterion movement length, number and length of interpolated movements, and the hand used in executing the criterion movement. The groups differed in the combination of hands used in executing criterion, interpolated, and recall movements. Both positive and negative directional biasing effects of interpolated movements were present for all groups, suggesting that interference indicative of the assimilation, or pooling, of criterion and interpolated memory traces was not specific to use of the same response mechanism for original and interpolated movements. The results indicated that relevance for the pooling of original and interpolated activity may depend on the similarity of the responses required, rather than on similarity or identity of response mechanisms, as suggested in 1970 by Pepper and Herman.

In a recent monograph, Pepper and Herman (1970) have offered a dual process theory of motor short-term memory (STM) to account for their own and prior research findings. Basically, they argue that both memory trace decay and interference processes interact in the retention of a motor act. The process of a memory trace for a criterion motor act is seen to be one of decay over time in the represented intensity or extent of the original response. This assumption, together with the evidence for "response sets," or tendencies to overshoot low amplitude or low intensity target responses and to undershoot high intensity responses, is seen to account for the conflicting evidence regarding recall error as a function of length of the retention interval.

In some cases, an increase in absolute error scores has accompanied an increase in the duration of a retention interval (Adams & Dijkstra, 1966; Posner, 1967; Posner & Konick, 1966), while Pepper and Herman (1970) found that error decreased with an increasing retention interval. In the former cases, as Pepper and Herman pointed out, a response set of undershooting is present, and hence, the error tends to be negative in sign. This combines with a decaying memory trace to produce an increasingly greater negative error and consequently a greater absolute error. However, in the cases where an overshoot response set results in a positive error, the decaying representation of the criterion actually yields more accurate recall responses, resulting either in less positive error or a small negative error. This, in turn, means that the absolute error will decrease during a retention interval.

The memory trace for a motor act is also seen by Pepper and Herman (1970) to be subject to modification by activity interpolated between criterion and recall responses. Evidence that interpolated nonmotor activities such as counting or digit classification tasks increase the amplitude (Posner, 1967) or intensity (Pepper & Herman, 1970) of the recall response suggests to them that such activities augment the memory trace, perhaps by increasing the general muscle tension, either directly or by distracting *S* from the primary task. When motor responses are interpolated and these involve essentially the same muscle groups as the criterion response, the momentary representation of the original response is again seen to be modified by a process of assimilation (Helson, 1964) of

---

[1] This research was supported by Grant GB6775 from the National Science Foundation.

[2] Requests for reprints should be sent to Don Trumbo, 544 Psychology Building, Pennsylvania State University, University Park, Pennsylvania 16802.

[3] This paper is based on a thesis by the second author, who is now with Merck, Sharpe, & Dohme, West Point, Pennsylvania.

the traces of the original and interpolated responses. That is, the trace of the original response is augmented when the interpolated specific motor acts are of greater intensity or amplitude than the original, but reduced when the interpolated acts are of lesser intensity or amplitude than the original. Since the recall response is assumed to be governed by the momentary representation of the original response, interpolated specific acts of greater magnitude than the original act should result in greater positive (or smaller negative) error at recall, while those of lesser magnitude should lead to smaller positive (or greater negative) errors, the sign of the error being dependent on the response set for the given criterion act.

Thus, two processes are postulated to interfere with the normal decay of the motor memory trace. The first, involving nonmotor activities, is seen to augment the memorial representation of the original movement, possibly by increasing the general muscle tension; the second, involving specific responses with the same muscle groups, either *augments* or *weakens* the representation through a process of assimilation, or pooling, of the original and interpolated memory traces. It would appear, then, that nonmotor interpolated activity can operate only to augment the representation, since any such activity would increase muscle tension relative to the inactive control condition, whereas responses which involve the same muscle groups as the criterion response can either augment or weaken the representation, depending on their intensity relative to the criterion response.

Although Pepper and Herman (1970) developed their theory from observations of the recall of pressure responses, their reanalysis of studies involving movement responses suggests that the theory may be extended to motor STM in general and merits further investigation. However, evidence relevant to the theory from studies of motor behavior other than pressure response is scant. Stelmach (1969a, 1969b, 1970) and Stelmach and Wilson (1970) have investigated the effects of both prior and interpolated positioning responses on the recall of a criterion position. While they found some evidence of both PI and RI effects, they report only scant evidence for the directional biasing predicted by the Pepper and Herman theory. They concluded that interpolated activity increases the variability of recall responses, rather than producing shifts in the mean. Even though Stelmach's studies are relevant to the question of the generality of the Pepper and Herman theory, they required distinctly different responses of the *S*s than the present study. Whereas Stelmach's *S*s are instructed to reproduce the terminal position of a movement response and movement length is made irrelevant by varying the starting point, the task in the present study is to reproduce the *length* of the criterion movement with terminal position an irrelevant cue. It is quite conceivable that the effects of interpolated activity are different for such different task requirements.

The purpose of the present study was both to examine some hypotheses from the Pepper and Herman theory (1970) and to test an extension of them. It seems reasonable to believe that interpolated motor responses which involve muscle groups other than those of the original response should have some effect. At the very least, they should produce an increase in general muscle tension which augments the trace of the original response, as appears to be the case with interpolated nonmotor responses. On the other hand, it is conceivable that information processing proceeds in a manner such that an interpolated response of the same class as the original response will have a similar effect whether or not the same muscle groups are employed in its execution. According to this line of reasoning, if, for example, both the original and the interpolated responses involve given extents of movement of a hand and arm, it should not greatly matter whether the responses are executed with the same or the other hand. Assimilation, or pooling, resulting in either an augmented or weakened memory trace would thus be predicted on the basis of response similarity rather than the specific muscle groups employed.

To this end, the present study investigated combinations of response mechanisms used in executing original, interpolated, and recall responses. In addition, the amplitude of original and interpolated responses, the number of interpolated responses, and the hand used for the original response were also studied.

## METHOD

*Subjects.*—The *S*s were 48 male and 32 female undergraduates who received research participation credit in the introductory psychology course for their services. All *S*s were right-handed.

*Apparatus.*—The apparatus was similar to that used by Adams and Dijkstra (1966). It consisted of a 12-mm.-diameter metal rod, approximately 1 m. in length, mounted between two vertical supports fastened to a wooden base. A freely moving slide was equipped with a finger grip handle and a pointer which moved along a meter stick fixed above and parallel to the rod. Thus, movement of the slide could be measured by referring to the meter stick. Also attached to the rod were two movable stops which could be fixed to determine the length of any controlled movement of the slide.

*Experimental design.*—Four independent groups of *S*s were used in a 4 × 4 × 2 × 2 × 2 design. The four groups, with 12 males and 8 females each, received different combinations of the same (S) and opposite (O) hands with respect to the criterion, interpolated, and recall movements. These combinations were: SSS, SOS, OSS, and SSO, in which, for example, SOS indicates that all criterion and recall movements were made with the same hand, while the interpolated movement(s) were made with the opposite hand.

The four within-*S* variables were (*a*) length of the criterion movement (Cl = 5, 10, 20, or 30 cm.); (*b*) hand with which the criterion movement was made (Ch = left or right); (*c*) length of the interpolated movement relative to the criterion movement (Il = longer or shorter); (*d*) number of repetitions of the interpolated movement (In = 1 or 5). The interpolated movements were 3 cm. and 7 cm., 5 cm. and 15 cm., 10 cm. and 30 cm., and 20 cm. and 40 cm., for the 5-, 10-, 20-, and 30-cm. criterion movements, respectively.

Each *S* received one trial on each combination of the within-*S* variables, or 32 experimental trials, plus 8 control trials, one for each of the combinations of hand (Ch) and length (Cl) of criterion movement, without interpolated activity. The experimental trials were randomized, first by generating a 4 × 4 Latin square wherein each of the four criterion movement lengths appeared once in each column and row, followed once and only once by each other value. This square was replicated to produce eight blocks of four trials, with values of each of the three two-level variables alternated within each block such that four *S*s within each independent group always received a control trial at the first position within each block, a second set of four *S*s always received a control trial at the second position in each block, and so forth. Finally, each *S* within these sets of four began the trials at a different block (i.e., row of the Latin square). The *S*s were randomly assigned both to groups and to trial orders within groups.

*Procedures.*—After a general description of the experiment and an opportunity to manipulate the apparatus, *S* was blindfolded and instructed to place his hands on the table. Before each movement, *E* placed *S*'s hand on the slide. At the command "move," *S* moved the slide from right to left to the stop which was set to determine the appropriate length of movement. At the command "return," *S* returned the slide to the original position against the right-hand stop. The *S* returned his hand to the table after each movement, except when five interpolated movements were required; all of these were made without releasing the slide. At the command "recall," *S* attempted to reproduce the criterion movement with the slide set at a new randomly determined starting position.

On all trials, the retention interval was 20 sec. and the intertrial interval, 30 sec. All recall movements were recorded to the nearest centimeter. The *S* was given 2 practice trials initially and a 2-min. rest midway through the 40 trials.

## RESULTS

The data were analyzed in terms of both absolute and algebraic error in reproducing the criterion distances at recall. However, the results are to be understood not only in terms of deviations from the criterion distances, but also relative to the response sets of the *S*s, i.e., tendencies to overshoot shorter criterion distances and to undershoot longer distances. For the present study, the magnitudes of these response sets are estimated by the data for the control trials wherein no interpolated movements were made. Mean absolute and algebraic errors for these trials are shown at the top of Table 1. The algebraic error data clearly indicate the magnitudes of these constant errors: the 5-cm. movement was overshot by about 1 cm., the 10-cm. movement showed essentially zero constant error, and the 20-cm. and 30-cm. movements were undershot by about 2.5 cm. and 4.0 cm., respectively.

Given these response set data as a base line, it was predicted that relatively long interpolated movements would increase error at 5 cm. and 10 cm., but decrease

TABLE 1

MEAN ABSOLUTE AND ALGEBRAIC ERRORS AT RECALL FOR ALL LEVELS OF ALL VARIABLES AS A FUNCTION OF CRITERION DISTANCE

| Variable | Length of the criterion movement (in cm.) | | | | |
|---|---|---|---|---|---|
| | 5 | 10 | 20 | 30 | $\bar{X}$ |
| Control (response set) | 1.48(1.02) | 1.64(−.04) | 3.84(−2.48) | 5.44(−4.30) | 3.09(−1.45) |
| Hand combination | | | | | |
| SSS | 1.55(1.15) | 1.92(.43) | 3.92(−.68) | 5.03(−2.01) | 3.10(−.28) |
| SOS | 1.29(.66) | 1.71(−.39) | 4.09(−1.40) | 5.48(−3.72) | 3.14(−1.21) |
| OSS | 1.51(.90) | 1.91(.19) | 4.05(−1.78) | 5.89(−3.57) | 3.34(−1.06) |
| SSO | 1.63(1.17) | 1.94(.12) | 4.03(−1.83) | 5.61(−3.26) | 3.30(−.95) |
| Criterion movement (hand) | | | | | |
| Right | 1.38(.83) | 1.87(.09) | 4.16(−1.59) | 5.76(−3.38) | 3.29(−1.01) |
| Left | 1.61(1.11) | 1.88(.09) | 3.88(−1.25) | 5.24(−2.90) | 3.15(−.74) |
| Interpolated movement (relative length) | | | | | |
| Longer | 1.84(1.58) | 1.87(.78) | 3.63(.05) | 4.86(−1.81) | 3.04(.15) |
| Shorter | 1.15(.37) | 1.88(−.60) | 4.42(−2.90) | 6.14(−4.47) | 3.39(−1.90) |
| Interpolated movements (number) | | | | | |
| One | 1.48(.95) | 1.81(.03) | 3.72(−1.77) | 5.53(−2.68) | 3.13(−.87) |
| Five | 1.51(.99) | 1.93(.15) | 4.33(−1.07) | 5.48(−3.59) | 3.31(−.88) |
| $\bar{X}$ (criterion movement, length) | 1.50(.97) | 1.85(.08) | 4.00(−1.50) | 5.49(−3.24) | 3.21(−.93) |

Note.—S refers to "same" hand and O to "other" hand. Thus, for example, SOS means that the same hand was used for the criterion and recall responses, while the other hand was used for interpolated movements. Algebraic errors are given in parentheses.

error at 20 cm. and 30 cm., while relatively short interpolated movements would decrease error at 5 cm., but increase error for 10-, 20-, and 30-cm. movements. These same effects should be amplified for five interpolated movements relative to that for only one.[4]

Table 1 also includes absolute and algebraic error means for all levels of each experimental variable as a function of the length of the criterion movement. The overall analyses of variance confirmed what these data suggest, namely, that only length of criterion movement and relative length of the interpolated movement(s) were significant main effects. The values for criterion length were $F\ (3,\ 228) = 219.4$ and 197.4, both $p\text{s} < .001$ for absolute and algebraic error, respectively; while for interpolated length, the corresponding values were $F\ (1,\ 76) = 6.9$, $p < .05$, and 178.2, $p < .001$.

The directional biasing effects are evident in a number of interactions in the analyses. These are presented graphically in Fig. 1 which shows absolute (A) and algebraic (B) error as a function of the number and relative length of interpolated movements. The significant Criterion Length × Interpolated Length interaction for absolute error, $F\ (3,\ 228) = 22.54$, $p < .001$, supports the prediction that shorter interpolated movements increase accuracy at shorter criterion distances, and vice versa. In addition, a significant Criterion Length × Interpolated Length × Interpolated Number interaction, $F\ (3,\ 228) = 4.48$, $p < .01$, indicated that this directional biasing was enhanced for five as compared with one interpolated movement.

The algebraic error data (see Fig. 1B) also indicate the directional biasing effects. In addition to the main effect of length of interpolated movements, the Interpolated Length × Interpolated Number and Interpolated Length × Criterion Length interactions were significant, with $F\ (1,\ 76) = 22.39$, $p < .001$, and $F\ (3,\ 228) = 9.98$, $p < .001$, respectively. The interaction of length and number of interpolated move-

---

[4] It is conceivable, though unlikely, that where decreases in error are predicted, the directional biasing effects would be of such magnitude as to produce errors opposite in sign and larger than those for the control condition.

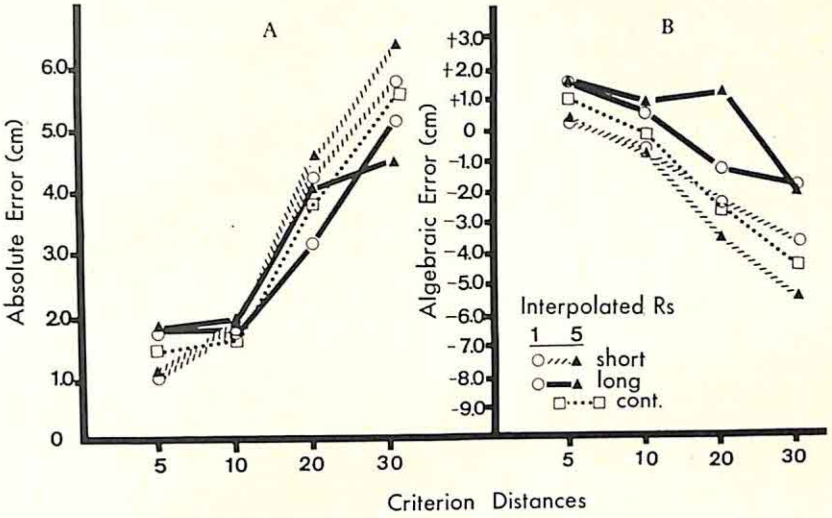


FIG. 1. Mean absolute (A) and algebraic (B) error in centimeters at each criterion distance for all combinations of length and number of interpolated movements.

ments indicates that the effect of number of movements was greater for the relatively long than for the shorter interpolated movements. This appears to be a result of the particularly large difference in error between one and five long movements at the 20-cm. distance. The Criterion Length × Interpolated Length interaction confirms the evidence of greater directional biasing at the longer criterion distances. Finally, significant interactions for Criterion Length × Interpolated Number, $F$ (3, 228) = 10.60, $p < .001$, and Criterion Length × Interpolated Length × Interpolated Number, $F$ (3, 228) = 4.82, $p < .01$, were indicative that the directional biasing effects were generally enhanced with five interpolated movements, especially at the longer criterion distances.

The net effects of the response sets were evident in Fig. 1B in the deviations of the control data from zero error, and the directional biasing effects of the interpolated movements are evident in the deviations of the data of the experimental conditions from these control data. However, the directional biasing effects are more clearly evident in Fig. 2, in which the data for the experimental conditions are shown as deviations from the control data, that is, with the control trials data set equal to zero. The effects of number of interpolated movements are less clear, indicating that, in general, five movements have a greater biasing effect than one movement, but only

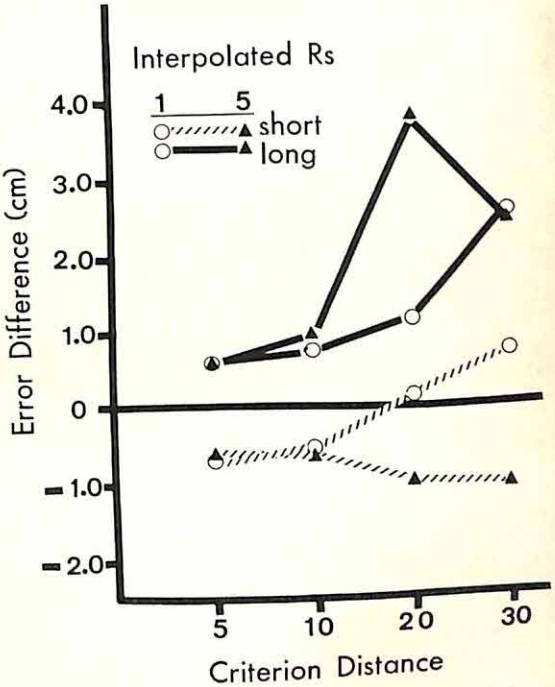


FIG. 2. Mean deviations (algebraic error) from the response set (control trials) data for all combinations of length and number of interpolated movements.

for the two longest criterion distances. These data depart from expectations at two further points: five long interpolated movements had no more effect than one movement at 30 cm. and one short interpolated movement was not associated with negative biasing at 20 or 30 cm.

Three additional interactions, Criterion Length × Criterion Hand and Criterion Length × Criterion Hand × Hand Combination for the absolute error and Criterion Hand × Hand Combination × Interpolated Length for the algebraic error were significant ($p < .05$), indicating that the between-*S*s variable of the hand combination used in executing the three classes of movements interacted with both the length and the hand used in executing the criterion movement. These results may reflect, in part, the fact that all movements were made from right to left and return, regardless of the hand used.

Of somewhat greater interest is the null finding with respect to the main effect of hand combination, indicating that the observed directional biasing effects were independent of the combination of response members used in executing criterion, interpolated, and recall movements. The interaction between hand combination and relative length of interpolated movement(s) resulted in a nonsignificant $F$ of .28, providing further evidence that the effects of interpolated movements were independent of the combination of response members employed.

The direct test of the hypothesis that the assimilation effects of interpolated motor response are not necessarily restricted to movements involving the same muscle groups is provided by the comparison between Cond. SSS and SOS in which criterion and recall movements were made with the same hand, but where, in the latter condition, interpolated movements were made with the opposite hand. The data for these two groups for each condition of interpolated movements at each criterion distance were highly similar, with each group showing both positive and negative biasing for long and short interpolated movements, respectively.

Comparison of the two remaining groups, SSO and OSS, in which criterion and recall movements were always made with opposite hands also failed to indicate any consistent differences.

A reanalysis of the data included the sex of *S*s as a sixth variable. Sex was found to be a reliable source of variance for the algebraic, $F$ (1, 72) = 5.34, $p < .05$, but not for the absolute error scores. However, separate analyses of variance showed that all the main effects and the interactions necessary to demonstrate the directional biasing effects were significant ($p < .05$) for both males and females. The sex difference was largely a difference in level, with females averaging $-$ 1.45 cm. error and males $-.49$ cm. error. The only other differences of interest were a significant "hand" effect, $F$ (1, 44) = 5.98, $p < .05$, and a significant Criterion Hand × Interpolated Number interaction, $F$ (1, 44) = 7.29, $p < .01$, for males, indicating that males were more accurate with the left than with the right hand and that for them the effects of number of interpolated movements was related to the hand used in executing the criterion movement.

## DISCUSSION

The response set, or range effect, data agree rather well with prior results (Adams & Dijkstra, 1966; Brown, Knauft, & Rosenbaum, 1948; Posner, 1967; Posner & Konick, 1966; Stelmach & Wilson, 1970), indicating that short movements are overshot and longer movements are undershot at recall.

The present data also provide confirmation of directional biasing effects, both positive and negative. The general effect of interpolating movements which were shorter than the criterion movement was to decrease the length of movements produced at recall, while interpolating relatively longer movements increased the length of recall movements. In absolute terms, these effects tended to increase with criterion distance for the positive biasing, but to remain fairly constant for the negative biasing. However, the *relative* amount of error attributable to biasing was roughly constant for longer interpolated movements and decreased with criterion distance for shorter interpolated movements. Increasing the number of interpolated movements had, in general, the expected effect of increasing the directional biasing, but this was not evident at the two shorter criterion distances nor for positive biasing at the longest criterion distance.

However, Pepper and Herman (1970) do not appear to be correct in their hypothesis that these bidirectional assimilation effects are dependent upon use of essentially the same response mechanism for both interpolated and criterion responses since the present results show both positive and negative directional biasing effects regardless of which hand and arm was employed in original and interpolated movements. Positive directional biasing for the various hand–arm combinations might be a result of either assimilation effects or of nonspecific activity by increasing muscle tension, since both of these should operate to augment memorial representation and result in positive biasing at recall. Evidence of negative directional biasing for all hand combinations, however, indicates rather clearly that assimilation effects are not dependent on use of the same muscle system in criterion and interpolated movements. In this case, only specific activity (i.e., involving the same muscle group) would be expected to result in increased negative error; any nonspecific activity should increase muscle tension relative to the control (rest) conditions and result in augmented memory traces. The absence of a significant hand combination main effect or a Hand Combination × Interpolated Length effect, together with clear-cut evidence for directional biasing, thus argue strongly that assimilation-like effects are occurring regardless of the combination of response mechanisms used.

It has been suggested that the data indicative of response biasing could have resulted from response competition, or confusion errors, where in some instances *S*s reproduced the interpolated rather than the criterion movement. This would have the effect of pulling the mean toward the length of the interpolated movement, yielding data consistent with that attributed to response biasing. If such errors occurred on the experimental trials, they should increase the intra-*S* variance for experimental as compared with control trials, since the latter should be free of confusion errors.

Unfortunately, the design, with a single observation per *S* per cell, precluded a test of this alternative explanation with the present data. However, the possibility that *S*s were confused as to which response to reproduce seemed worthy of further testing. Consequently, in a follow-up experiment, six *S*s were tested under conditions similar to those in the original study, except that they received only four types of trials: control trials with either 20- or 30-cm. criterion distances and experimental trials with either 20- or 30-cm. distances, and 30- and 20-cm. interpolated movements, respectively. Each trial type was presented 10 times so that intra-*S* variances could be assessed. The 40 trials were presented in 10 random orders of four events, and all recall responses were called for after a 20-sec. retention interval.

Of primary interest were the comparisons of intra-*S* variances on control and experimental trials for the same criterion distance and the extent of overlap between distributions of responses for the two different criterion distances.

The results provided no support for the notion of response competition or confusion errors: of the 12 comparisons between response distributions, intra-*S* variability was greater in six instances for control trials and in six instances for experimental trials. Furthermore, in only 3 of the 12 comparisons was a response to one criterion distance found to overlap the distribution of responses to the other criterion distance. In each of these cases, only a single observation overlapped and one of these occurred in the control data, effectively canceling one instance in the experimental data. The single remaining instance was 1.45 standard deviations from the mean of the overlapped distribution, i.e., in the extreme 7% of the distribution.

While these data do not demonstrate conclusively that response competition or confusion errors did not occur in the original study, they provide no support for such an interpretation. Furthermore, while neither the original nor the follow-up results provide conclusive evidence that directional biasing effects result from the pooling of criterion and interpolated memory traces, they are quite consistent with such an interpretation, providing one does not interpret "pooling" to mean a simple arithmetic averaging of the two memorial events. Relatively long interpolated movements appear to result in an augmented memorial representation of the criterion movement at recall, while relatively short interpolated movements result in a reduced representation at recall. However, present findings show no indication that original and interpolated responses are weighted equally in the memory trace governing the recall response. A recent study by Patrick (1971) indicates that the recency of the interpolated response may determine in part the extent to which it biases the recall response:

he found that a single interpolated response executed late in the retention interval had as much effect as five interpolated responses executed throughout the interval. In addition to indicating one parameter of directional biasing and possibly of the mechanism underlying this phenomenon, Patrick's results may also account for the limited effects of number of interpolated responses in the present results. For the present, the term "pooling" must be taken as descriptive of the directional biasing effects of interpolated activity. Its value as an explanatory concept must be determined by further research.

In contrast with the position of Pepper and Herman (1970), the present data suggest that responses which are similar in *class* will be encoded and stored in the same way, regardless of the specific muscle groups employed in responding, and that under these conditions bidirectional assimilation-like effects will occur. Response similarity in this sense appears to be more closely related than identity of peripheral response mechanisms to the concept of relevance in adaptation level theory (Helson, 1964) as well as to the concept of similarity in the literature on memory.

## REFERENCES

Adams, J. A., & Dijkstra, S. Short-term memory for motor responses. *Journal of Experimental Psychology*, 1966, **71**, 314–318.

Brown, J. S., Knauft, E. B., & Rosenbaum, G. The accuracy of positioning responses as a function of their direction and extent. *American Journal of Psychology*, 1948, **61**, 167–182.

Helson, H. *Adaptation-level theory.* New York: Harper & Row, 1964.

Patrick, J. P. The effect of interpolated motor activities in Short-term motor memory. *Journal of Motor Behavior*, 1971, **3**, 39–48.

Pepper, R. L., & Herman, L. M. Decay and interference effects in the short-term retention of a discrete motor act. *Journal of Experimental Psychology*, 1970, **83** (2, Pt. 2).

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, 1967, **75**, 103–107.

Posner, M. I., & Konick, A. F. On the role of interference in short-term retention. *Journal of Experimental Psychology*, 1966, **72**, 221–231.

Stelmach, G. E. Prior positioning responses as a factor in short-term retention of a simple motor task. *Journal of Experimental Psychology*, 1969, **81**, 523–526. (a)

Stelmach, G. E. Short-term motor retention as a function of response similarity. *Journal of Motor Behavior*, 1969, **1**, 37–44. (b)

Stelmach, G. E. Kinesthetic recall and information reduction activity. *Journal of Motor Behavior*, 1970, **2**, 183–194.

Stelmach, G. E., & Wilson, M. Kinesthetic retention, movement extent, and information processing. *Journal of Experimental Psychology*, 1970, **85**, 425–430.

(Received September 15, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 213–216

# VERIDICAL ROTATION IN DEPTH IN UNIDIMENSIONAL POLAR PROJECTIONS DEVOID OF THREE MOTION-PARALLAX CUES[1]

WAYNE HERSHBERGER[2] AND DAVID L. CARPENTER

*Northern Illinois University*

Observers, viewing a horizontal row of dots upon a frontal screen converging to and diverging from an end dot in a manner faithfully simulating rotation in depth, judged the direction of that simulated rotation. Although they saw only a pie-shaped sector of the circuit of rotation devoid of previously identified perceptual cues to rotation direction, *O*s identified rotation direction at a level significantly beyond chance. It is suggested that the acceleration/displacement disproportionalities of the various dots, or rather the stimulus gradient of these disproportionalities, mediates these veridical judgments.

Observers viewing, upon a frontal screen, a row of moving dots converging to and diverging from the center dot in a manner faithfully simulating rotation in depth (see Fig. 1) are able to identify the direction of that rotation by using each of three motion parallax cues (Hershberger & Urban, 1970): (*a*) the retinal direction (Direction) in which the motion of the projected dots on opposite ends of the line are periodically congruent (when the dotted line is near the frontal plane, all dots move in the same direction across the visual field: to the right during clockwise and to the left during counterclockwise rotation), (*b*) the order (Order) in which the dots reach their respective limits of angular displacement in the visual field (to the right of the axis, the outermost dot reaches its limit first during counterclockwise rotation and last during clockwise rotation; to the left of the axis the opposite is true), and (*c*) the differential retinal velocity (Velocity) of dots on the near versus far side of the axis of rotation (traditional "motion parallax"). The purpose of the present study was to determine whether these three cues may be the necessary as well as sufficient mediators of veridical perception of direction of rotation in depth in such a display.

It is possible to divide the rotary motion depicted in Fig. 1 into several parts, none of which incorporates any of the above cues. A partial motion projection simulating the moving dots only when they are momentarily beyond the axis of rotation (i.e., falling within Arc F R′ L′ F′) incorporates none of the above cues, nor does a partial projection simulating the moving dots only when they fall within the Sector LAR. The ability of *O*s viewing these partial projections to correctly identify direction of rotation would reveal the efficacy of yet another cue. On the other hand, the inability to do so would suggest

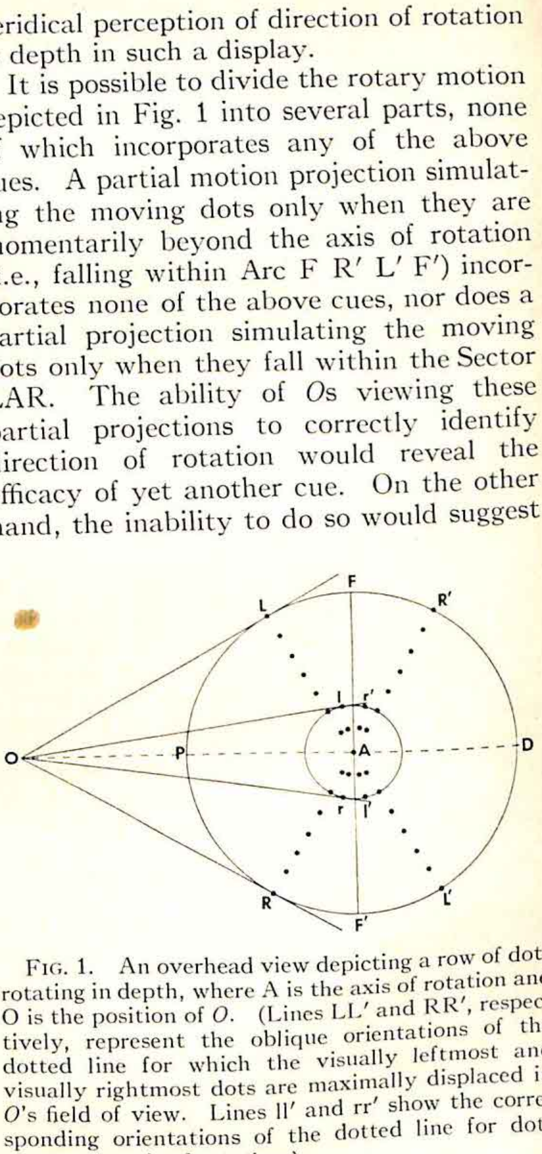


FIG. 1. An overhead view depicting a row of dots rotating in depth, where A is the axis of rotation and O is the position of *O*. (Lines LL′ and RR′, respectively, represent the oblique orientations of the dotted line for which the visually leftmost and visually rightmost dots are maximally displaced in *O*'s field of view. Lines ll′ and rr′ show the corresponding orientations of the dotted line for dots nearer the axis of rotation.)

---

[1] This investigation was supported by United States Public Health Service Research Grant 1 RO3 MH 20607-01 MSM from the National Institute of Mental Health. The authors thank Fred Attneave for suggesting the use of the parallel control projections and James Starzec for his assistance in testing *O*s.

[2] Requests for reprints should be sent to Wayne Hershberber, Psychology Department, Northern Illinois University, DeKalb, Illinois 60115.

that the three cues described above are the necessary as well as sufficient mediators of veridical perception of direction of rotation in such displays.

## METHOD

*Observers.*—Twenty-four college students enrolled in several psychology courses earned bonus course credit for participating as *O*s.

*Apparatus.*—The *O*s viewed through a narrow horizontal slot a vertically moving belt upon which were drawn sets of long curves lines. As the belt passed by the aperture, the visible dot-like segments of these lines moved along the horizontal slot simulating a dotted line rotating in depth about a vertical axis. The curved lines were drawn in black ink with a 1-mm. (diameter) pen point on a belt of translucent vellum paper 30.5 cm. wide and approximately 15 m. long. The drawing was done by an automated *XY* plotter from data points generated by a computer program. In generating the data points, the projection plane (FF′ in Fig. 1) was taken as coincident with the axis of rotation and normal to a line joining the axis (A in Fig. 1) and the projection point (O in Fig. 1). Six different projections were generated, four polar projections using a projection distance of 30.5 cm. and two parallel projections using an infinite projection distance. The parallel projections incorporate no stimulus information as to rotation direction and served as control projections for assessing response bias. Each of the projections was a partial projection in the sense that it represented only those parts of a rotating, dotted-line figure momentarily falling within a particular sector of the circuit of rotation. The two parallel projections, Parallel left and Parallel right, are represented in terms of Fig. 1 as only those parts of a rotating line figure momentarily falling within Sectors FAD and F′AD, respectively. The figure being represented was a cross comprised of two 22.9-cm. lines of 13 evenly spaced dots oriented orthogonally to each other and to the axis of rotation, which served as their common center. Both of these projections simulated two full revolutions of the cross beginning with the cross arms coincident with *O*'s sagittal-frontal axes (FF′ by OD in Fig. 1). One pair of polar projections (Polar Far Sectors: left and right) was simply the polar analogue of the corresponding control projections. The other pair (Polar Near Sectors: left and right) represented only those parts of a rotating figure momentarily falling within 45° sectors extending respectively clockwise and counterclockwise from Line PA in Fig. 1. The figure being represented was a radially symmetrical arrangement of four 22.9-cm lines of 13 evenly spaced dots crossing at their center so as to resemble an eight-armed pinwheel. Each of these two projections simulated one full revolution of this pinwheel beginning with one of its four lines coincident with *O*'s frontal plane (FF′ in Fig. 1). By using the four- and eight-armed figures for the 90° and 45° sector projections, respectively, one and only one of the arms fell at any given time into the sector being projected. In turn, *O*, in viewing any one of the six partial projections, viewed a sequence of eight, successive, single straight lines of seven dots each rotating in depth about one end dot situated in the middle of the viewing slot. And, for the first and second members of each pair of projections, the dots appeared respectively to *O*'s left and to *O*'s right of the axis of rotation. Since it was not possible to match the three types of projections (Parallel, Polar Far Sector, and Polar Near Sector) with respect both to retinal and rotational velocity, it was decided to match them on retinal velocity. This was done by adjusting the length of the projections as they were drawn upon the paper belt. Since the displacement of the projected dots for a 45° rotation in the Polar-Near-Sector projections was comparable to that for a 90° rotation in the Polar-Far-Sector and Parallel projections, the retinal velocity was equated by halving the rotational speed of the Polar-Near-Sector projections. This was done simply by drawing one revolution of these near-sector projections as long as two revolutions of the other two types of partial projections: 203.2 cm.

Once the projections had been recorded, the paper belt was mounted as a scroll onto two horizontal spools, one mounted above the other, inside a large black metal box. The belt, which could be reeled from one spool to the other either up or down at a constant speed of 304.8 cm/min passed immediately behind one side of the enclosure in which was cut a horizontal slot .4 mm. high × 29.2 cm. long. The *O* viewed the stimulus displays via this slot. A 30.5-cm. long, 14-w. fluorescent tube, aligned parallel to and 1.27 cm. behind the slot, illuminated the paper belt from the rear.

*Procedure.*—Each *O*, tested individually, observed all six projections monocularly from a distance of 61 cm. Each of the three types of projections was presented to *O* eight consecutive times, four using the right-sector projection and four using the left-sector projection, with two of each of these trials simulating clockwise and the other two counterclockwise rotation. Half of the *O*s viewed them in the counterbalanced order Lc, Lcc, Rcc, Rc, Rcc, Rc, Lc, Lcc and half in the order Rcc, Rc, Lc, Lcc, Lc, Lcc, Rcc,Rc. The three blocks of trials, one block for each of the three pairs of projections, were presented in all possible orders (six) with four *O*s assigned randomly to each.

During each trial, *O* used a rotary swtich to continuously indicate the direction of apparent rotation. He released the switch only when he was uncertain. Each trial lasted 40 sec. The intertrial interval was on the order of 5 sec. and the interblock interval about 90 sec.

## RESULTS AND DISCUSSION

The means and standard deviations of the durations (in seconds) of clockwise and counterclockwise judgments for leftward and rightward retinal motions are shown in Table 1 for each of the three partial projections: Parallel, Polar Far Sector, and Polar Near Sector. The table includes means both for the raw data and the raw data corrected for response bias.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE DURATION (IN SECONDS) OF CLOCKWISE AND COUNTERCLOCKWISE JUDGMENTS FOR LEFTWARD AND RIGHTWARD RETINAL MOTIONS FOR EACH OF THREE PROJECTIONS: PARALLEL, POLAR FAR SECTOR, AND POLAR NEAR SECTOR

| Direction of retinal (and rotational) motion | Judged direction of motion | | | |
|---|---|---|---|---|
| | Clockwise | | Counter-clockwise | |
| Parallel projection | | | | |
| Right to left | 8.1<br>(8.7) | | 25.0<br>(11.4) | |
| Left to right | 22.2<br>(9.9) | | 11.9<br>(9.1) | |
| Polar-Far-Sector projection | | | | |
| Right to left (clockwise) | 13.5<br>(8.4) | *5.4*<br>(*8.4*) | 20.7<br>(11.9) | *−4.3*<br>(*11.9*) |
| Left to right (counter-clockwise) | 16.2<br>(9.9) | *−6.0*<br>(*9.9*) | 16.9<br>(10.1) | *5.0*<br>(*10.1*) |
| Polar-Near-Sector projection | | | | |
| Left to right (clockwise) | 27.1<br>(8.04) | *4.9*<br>(*8.04*) | 7.9<br>(8.0) | *−3.0*<br>(*8.0*) |
| Right to left (counter-clockwise) | 5.1<br>(6.4) | *−3.0*<br>(*6.4*) | 28.5<br>(8.1) | *3.5*<br>(*8.1*) |

Note.—Standard deviations are shown in parentheses. The means for the parallel projection are taken as a measure of response bias, which when subtracted from the corresponding means for the polar projections yield the corrected means printed in italics. All the resulting residuals deviate from response bias in a veridical direction; i.e., an illusory residual of negative duration represents the duration by which illusory response bias was reduced or offset by a tendency for veridical perception of rotation direction.

In the partial polar projections used here, direction of simulated rotation is correlated with direction of retinal motion: for the far-sector projections, clockwise rotation is always right to left motion across the stationary retina (left to right across the visual field); and counterclockwise rotation is left to right retinal motion. For the near-sector projections the reverse is true. Consequently, any response bias or tendency on the part of *O* to interpret a particular retinal direction as signifying a particular rotational direction would artificially inflate the veridicality of *O*'s judgments of either the near-sector or far-sector polar projections, while deflating the veridicality of his judgment of the other. Further, since these effects though opposite are not necessarily equal they cannot be expected to cancel one another in any combination of the near-sector and far-sector judgments. It is necessary to actually measure response bias and subtract it from the data. The parallel projections were used for this purpose. Although they incorporated no information as to direction of rotation, they did yield an impression of rotation in depth (cf. Braunstein, 1966) such that *O*s found little difficulty in making judgments of apparent rotation. For a trial of 40-sec. duration, the mean judged durations of apparent clockwise and counterclockwise rotation yielded by the parallel projections were, respectively, 8.1 and 25.0 sec. for leftward *retinal* motion (rightward in the visual field) and 22.2 and 11.9 sec. for rightward *retinal* motion. It is as if *O* is predisposed to see the axis of rotation as being situated on the far end of the rotating line. This response bias artifically inflates and deflates, respectively, the veridicality of the near-sector and far-sector judgments. To control for this response bias, four means of the type just described were computed for each *O* and in turn subtracted from his corresponding means for the two types of polar projections. Using this corrected data, the difference between mean duration of veridical and illusory judgments was computed for each *O* for both the near-sector and far-sector polar projections. If *O*s were guessing, these difference scores for both types of projection should equal zero on the average. They do not. For the far-sector and near-sector projections, the mean differences were 10.4 sec. ($SD = 11.9$) and 7.2 sec. ($SD = 12.1$), respectively. Both are significantly greater than zero, $t\ (23) = 4.28$, $p < .001$, and $t\ (23) = 2.90$, $p < .01$, respectively. The two means are not significantly different from one another, $t\ (46) = .92$, $p > .05$.

What is the nature of the stimulus mediating this veridicality? What asymmetry of stimulation does a polar projection of these near and far sectors of the circuit of rotation provide which a parallel projection does not? In a parallel projection of a dotted line rotating in depth, the projected displacement of a dot unit distance from the axis of rotation equals Sin $\theta$, where $\theta$ is the angle of rotation of the dotted line about a line normal to the projection plane at the axis of rotation. Further, the acceleration of that projected dot, $d^2(\text{Sin}\ \theta)/d\theta^2$, is $-\text{Sin}\ \theta$. Hence, for any dot on the rotating line the acceleration (absolute magnitude) of that projected dot is directly proportional to its displacement (absolute magnitude), with that proportion being unity: $1/1 = 1$. This is not true of a polar projection. In a polar projection, the acceleration of a dot, relative to its displacement, is disproportionately great and small, respectively, in near and far sectors of the circuit of rotation. Further, this asymmetry increases as the radius of the circuit increases; i.e., for dots farther from the axis of rotation the asymmetry is greater. Hence, for a dotted line momentarily falling in a near sector (e.g., Arc LPR of Fig. 1) the acceleration/displacement (A/D) disproportionalities of the respective projected dots increase in magnitude in order of the increasing displacement of the projected dots from the common axis of rotation. In the far sector, the relationship between these two orders is reversed; i.e., as displacement increases the A/D disproportionalities decrease in magnitude. In other words, considering both ends of the dotted line at the same time, the projected dots are monotonically ordered in terms of their respective A/D disproportionalities such that the largest and smallest disproportionalities are those associated with the two end dots which are respectively nearest to and farthest from the *projection point*, i.e., the simulated position of the viewer. This asymmetry, absent from parallel projections, is a veridical cue to relative depth, which when coupled with the retinal direction of the respective dots, allows *O*, to identify rotation direction. That is, for dots seen beyond the axis of rotation, rightward and leftward *retinal* motions signify counterclockwise and clockwise rotation, respectively, whereas for dots on the near side of the axis of rotation the rule is reversed.

The results of the present investigation of rotation in depth in unidimensional displays require that an additional cue to rotation direction (at least one) be added to the previously identified list of three (Hershberger & Urban, 1970). The asymmetry of the A/D disproportionalities provides the requisite stimulus information and as such ranks as a likely candidate for the role.

## REFERENCES

BRAUNSTEIN, M. L. Sensitivity of the observer to transformations of the visual field. *Journal of Experimental Psychology*, 1966, **72**, 683–689.

HERSHBERGER, W. A., & URBAN, D. Three motion-parallax cues in one-dimensional polar projections of rotation in depth. *Journal of Experimental Psychology*, 1970, **86**, 380–383.

(Received October, 18 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 1, 217–218

## SHORT REPORT

# TRUTH, AMPHIGORY, AND THE SEMANTIC INTERPRETATION OF SENTENCES[1]

DANNY D. STEINBERG[2]
*University of Hawaii*

The truth-value interpretation of amphigorous ("meaningless"), contradictory, redundant, and informative sentences and their corresponding negations were investigated. Thirty *S*s rated 156 affirmatives and their 156 negatives according to whether they were true, false, either true or false, or neither true nor false. The same 312 sentences were also rated with regard to whether they were amphigorous, redundant, contradictory, or informative. Principally, the findings strongly support A. C. Ewing's and T. Drange's positions that amphigorous affirmatives, e.g., "The chair is a sheep," are interpreted as false while their negations, e.g., "The chair is not a sheep," are interpreted as true. A. Pap's thesis that all such sentences would be regarded as neither true nor false was not confirmed.

Pap (1960) has stated that sentences like "Socrates is a prime number" would "unhesitatingly be dismissed as meaningless, not false, by plain people [p. 41]" and that their negations, "Socrates is not a prime number," would also be regarded as meaningless, i.e., neither true nor false. On the other hand, Ewing (1965) and Drange (1965) interpret the affirmatives as false and their negations as true. The present study investigates the truth-value interpretation of such "meaningless" or amphigorous (A) sentences and their negations, and of redundant (R), contradictory (C), and informative (I) sentences and their negations as well. The research also attempts to replicate the findings of previous studies (Steinberg 1970a, 1970b) with respect to the sentence categories, R, C, I, and A.

*Method.*—The *S*s were 30 male and female college students ranging in age from 17 to 23 yr. All were native speakers of English.

Each of the 30 *S*s rated two booklets of sentences. Each booklet contained the same 312 sentences but the rating instructions for the two booklets differed. In one booklet, the TRUTH form, *S*s were required to rate sentences as (*a*) True (T), (*b*) False (F), (*c*) Either True or False (E), or (*d*) Neither True nor False (N). No examples were given. In the other booklet, the INFO form, *S*s were required to rate sentences on the basis of examples and the following category labels: (*a*) Redundant (R), (*b*) Contradictory (C), (*c*) Acceptable and Informative (I), and (*d*) Irrelevant, unacceptable, and not contradictory (hereafter called Amphigorous-A). The *S*s were given the INFO form first for rating. The task was completed in 2 hr. or less.

The sentences presented to *S*s for rating were constructed using two frames, "The ____ is a ____." and "The ____ is not a ____." and the lexical items MAN, WOMAN, PERSON, HUSBAND, WIFE, SPOUSE, FIANCE, FIANCEE, BETROTHED, RAM, EWE, SHEEP, and CHAIR.

Materials and sentence construction procedure are discussed in detail in Steinberg (1970a, 1970b).

*Results and discussion.*—On the basis of that category which received the highest frequency of response, i.e., the mode, each of the 312 sentences of the TRUTH form was designated as characteristically T, F, E, or N, and each of the 312 sentences of the INFO form was designated as R, C, I, or A. Thus, each sentence received two characteristic designations, one from the TRUTH form and one from the INFO form.

Table 1 presents the results, on the basis of positive and negative sentences for the TRUTH form alone (left side), the INFO form alone (center), and for the relations between the two forms (right side). The relation measure is based on the two ratings each *S* gave to each sentence. On the TRUTH form,

TABLE I

RATING RESULTS FOR THE SEPARATE TRUTH AND INFO FORMS AND FOR RELATIONS BETWEEN FORMS FOR POSITIVE AND NEGATIVE SENTENCES

| Sentence group | | TRUTH form | | INFO form | | TRUTH and INFO forms | |
|---|---|---|---|---|---|---|---|
| | | Category | % | Category | % | Relation | % |
| 1 | | | | | | | |
| | pos | T | 89.5 | R | 75.6 | T&R | 70.9 |
| | neg | F | 88.7 | C | 71.3 | F&C | 66.0 |
| 2 | | | | | | | |
| | pos | E | 82.0 | I | 86.0 | E&I | 74.0 |
| | neg | E | 83.3 | I | 82.8 | E&I | 73.2 |
| 3 | | | | | | | |
| | pos | F | 83.7 | C | 75.8 | F&C | 69.4 |
| | neg | T | 82.5 | R | 63.8 | T&R | 56.0 |
| 4 | | | | | | | |
| | pos | F | 77.0 | A | 65.3 | F&A | 53.3 |
| | neg | T | 86.2 | R | 46.9 | T&R | 42.5 |

Note.—T = Truth, F = False, E = Either T or F, R = Redundant, C = Contradictory, I = Informative, A = Amphigorous.

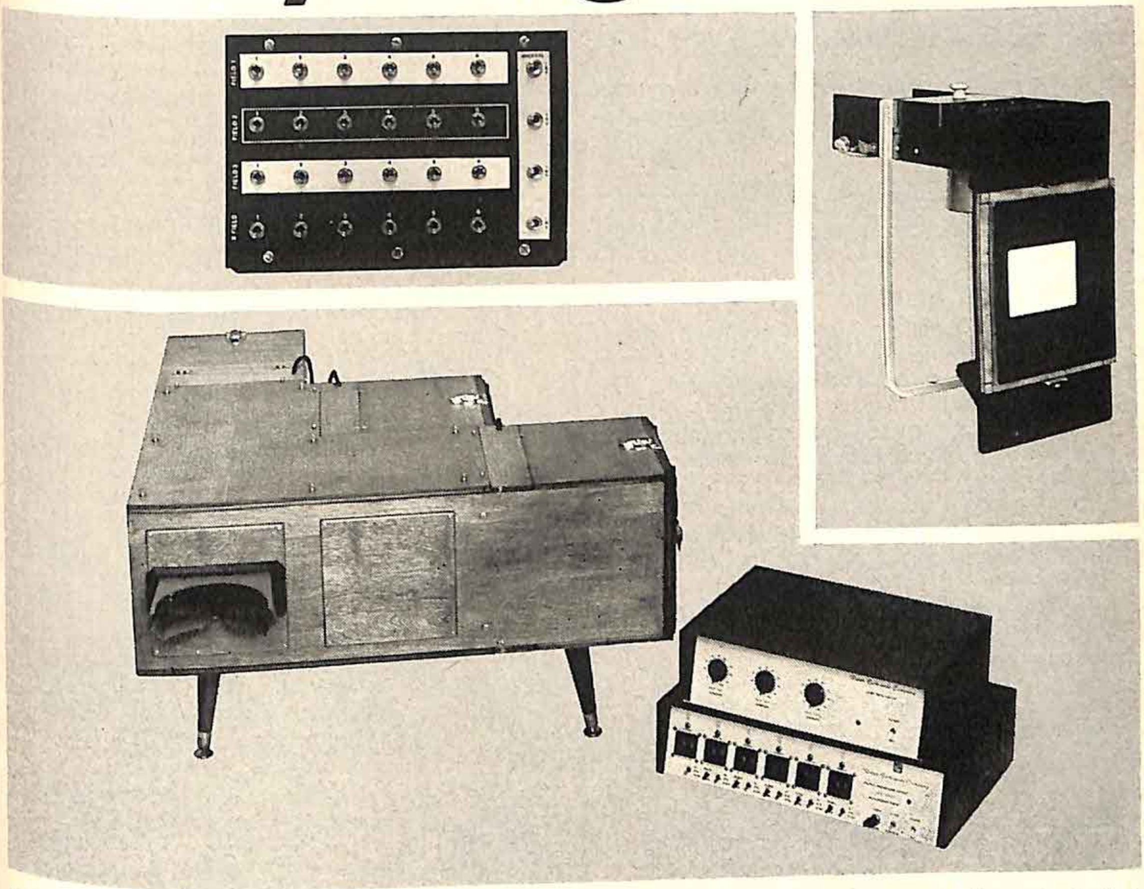
---

[1] This investigation was supported in part by a Public Health Service Fellowship, 1-F2-MH-35, 206-01(MTLH) from the National Institute of Mental Health and by a Biomedical Sciences grant from the University Research Council of the University of Hawaii.

I am greatly indebted to William May and to Sandra Winslow of the University of Illinois for their assistance in the analysis of the data.

[2] Requests for reprints should be sent to Danny D. Steinberg, Department of English as a Second Language, University of Hawaii, Honolulu, Hawaii 96822.

# at last ! easy alignment



That's right! For ease of alignment, all our tachistoscopes now have a new *center pivoted, gimbal mount* mirror system.

What else? Well, we have a new Universal Tachistoscope Logic Interface which enables you to program complex tachistoscopic functions. We also have a new Automatic Card Changer which attaches directly to the tachistoscope cabinet.

And, don't forget. All our Harvard Tachistoscopes come complete with direct reading electronic digital timers [accuracy ±.005%], solid-state lamp drivers, and interchangeable stimulus backs.

Versatile, reliable, and easy to align. Write for complete information.

Ralph Gerbrands Company
8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

for example, 89.5% of all responses to the positive sentences labeled as Group 1 were true; while on the INFO form, 75.6% of all responses to that same group of sentences were redundant. Of all the *S*s' responses, 70.9% were true on the TRUTH form and redundant on the INFO form, i.e., a T&I relation.

For the 156 different positive sentences four distinct relations obtained between forms. They are (*a*) T&R, e.g., "The ram is a sheep," (*b*) E&I, e.g., "The sheep is a ram," (*c*) F&C, e.g., "The man is a wife," and (*d*) F&A, e.g., "The chair is a man." When the four groups of sentences were negated, the relations became F&C, E&I, T&R, and T&(R or A), respectively.

The results clearly support the position of Drange (1965) and Ewing (1965) since positive amphigorous sentences such as "The chair is a sheep" were rated as False while their negations were rated as True. Pap's (1960) thesis finds no empirical support for there was not a single sentence which was rated as a majority of the *S*s as Neither True nor False. With regard to the replication aspect of this study, the results show that all of the sentences of the INFO form were categorized in exactly the same way as they were in the previous studies.

## REFERENCES

Drange, T. *Type crossings.* The Hague: Mouton, 1965.

Ewing, A. C. Meaninglessness. In P. Edwards & A. Pap (Eds.), *A modern introduction to philosophy.* New York: The Free Press, 1965, 705–714.

Pap, A. Types and meaninglessness, *Mind*, 1960, 41–54.

Steinberg, D. D. Analyticity, amphigory and semantic interpretation of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 37–51. (a)

Steinberg, D. D. Negation, analyticity, amphigory, and the semantic interpretation of sentences. *Journal of Experimental Psychology*, 1970. **84**, 417–423. (b)

(Received July 1, 1970)

# IT'S HERE!

# a human test system with professional features you can appreciate...

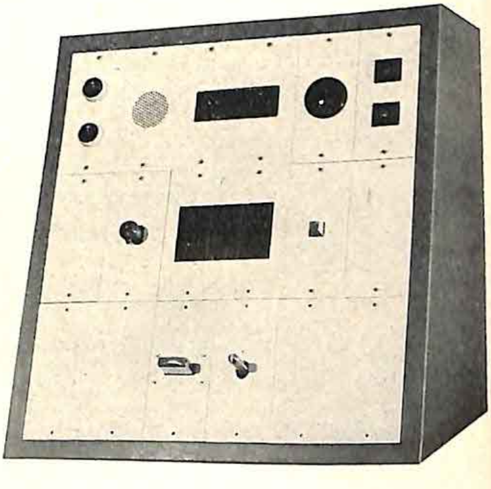
BRS-FORINGER has just recently introduced a line of Modular Human Test Systems into its ever expanding line of professional equipment. Laboratory Systems utilizing BRS-FORINGER Human Test Equipment are available for all types of applications including:

*BEHAVIOR MODIFICATIONS*
*SENSORY ASSESSMENT*
*OPERANT SCHEDULES*
*DISCRIMINATION*
*NONSENSE SYLLABLES*

*CONCEPT FORMATION*
*MATCH-TO-SAMPLE*
*AUDIO*
*ODDITY*
*ETC., ETC.*

Systems for specific applications are designed free of charge utilizing stimulus, response, and reinforcement console modules to meet the specific research or instructional requirements. The new BRS-FORINGER Human Test Systems incorporate many innovative, professional features that you can appreciate.

- *INTELLIGENCE MODULES - A complete line of modules designed to fit all applications. Attached to panels with flush fitting, self-clinching fasteners for tamper-proof construction.*
- *INTELLIGENCE PANELS - 1/8" aluminum to withstand typical and atypical punishment. Standard panel measures only 3-1/8" wide allowing for more intelligence modules per console.*
- *MODULE FINISH - Attractive yet extremely durable white baked epoxy finish.*
- *INTELLIGENCE PANEL MOUNTING CONNECTORS - "Camloc" structural fasteners utilizing tamper-proof 1/8" Allen type heads which set flush with panel when tightened.*
- *MOUNTING RAILS - For extra strength in construction, rails are .093" channel steel. Permanently welded to console chassis.*
- *CONSOLE - 1/16" thick steel. Models offered to accommodate two or three rows of intelligence panels.*



*Call TOLL FREE for further information and prices*
800/638-0858*

*Outside Continental U.S. and in the state of Maryland call us collect at: 301/474-3400

BRS FORINGER® 5451 Holland Drive • Beltsville, Maryland 20705
A DIVISION OF TECH SERV, INC.

# Audition and Instrumentation

## ...how to make your work more precise and your lab more flexible.

ICONIX equipment for your audition work: Our Audio Switch switches stimuli without masking distortion or transient cueing; Timer/Control Units select presentation order, measure reaction times and time intervals with repeatability and accuracy; our Logic Cabinet with Program Plugboard enables you to control most experimental parameters from one instrument with a wide range of built-in logic functions. You can establish your experimental program (easily altered as you wish), change to another program, then return to your original program quickly, economically and with unfailing repeatability.

Some features of our products:

**Audio Switch:** Model 0137, Single Channel – $180.

1. Frequency response to DC. Low frequency components of your stimuli will be undistorted. At 36 kHz, it's −3 dB.
2. Noise level <1 mv (−60 dB from 1 volt input), with an on/off ratio of better than 80 dB.
3. Gain is variable up to +6 dB.
4. Imperceptible switching transients. Total harmonic distortion <1.5% at full 1.5 v p-p input.
5. Adjustable rise/fall times: Fast time −15 μs, or a range of 1–250 milliseconds.

**Timer/Controller:** Model 6246 – $950. Model 6255 Timer with 6010 Controller (shown) – $1950.

Both options feature:

1. Crystal controlled accuracy of .001%.
2. Drift of .001% per YEAR . . . for repeatibility.
3. Intervals set by thumb wheel switches.
4. Front panel or remote operation.
5. Display of reaction time on front panel.
6. Intervals presented one time, or recycled in elapsed time mode. 6010 also has sequential mode.
7. Outputs for function control and data acquisition.

The 6255 Timer has ranges from 0.01 milliseconds to 1000 seconds, and the 6010 Controller sets 10 intervals (to five digits). The 6246 Timer/Controller has ranges from 0.1 milliseconds to 10 seconds, and controls six intervals (to four digits).

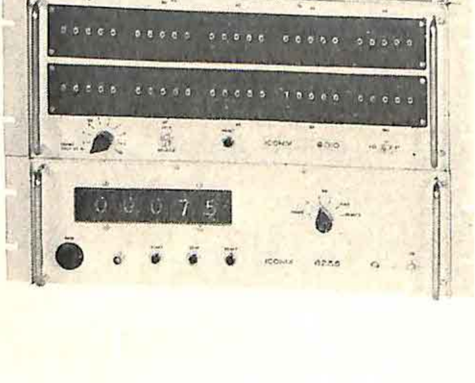
**Logic Cabinet with Program Plugboard:** Model 6171 – $500.

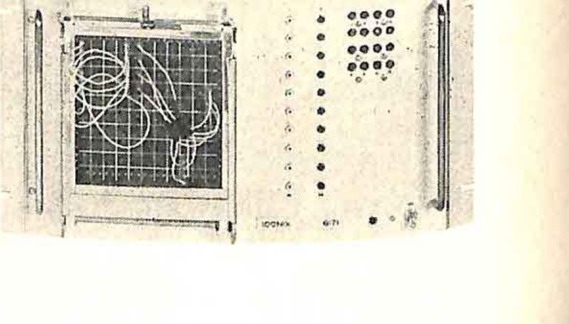
1. Cabinet space for 28 optional logic circuit boards; including logic gates, level shifters, memories, amps, pulse formers and relay drivers. Over 30 to choose from.
2. Power supplies included for logic level and relay operation can be used to power other equipment.
3. 440 point plugboard includes 100 patch cords.
4. 16 indicator lamps on front panel are wired to plugboard points for logic function display.
5. 20 front panel pin jacks and 40 rear panel jacks wired to plugboard points, for your systems use.

There's more: Our new line of analog products such as a programmed attenuator, mixers and power amplifiers. Write or phone to learn about our standard line of more than 30 proven and reliable products, our unique warranty (we guarantee our specifications) and our "hot line" service policy. We'll send our booklet showing how we can make your research easier . . . and a lot less expensive.

*Sales Service and Applications Assistance:*
*(408) 243-2424*

Siliconix incorporated
2201 Laurelwood Road · Santa Clara · CA 95054
Telephone (408) 246-8000 · TWX: 910-338-0227





# *NOW AVAILABLE*

## *A SOURCE BOOK FOR THE FIRST COURSE IN PSYCHOLOGY*

A source book for teachers of general psychology, containing reviews of introductory texts, books of readings, and lab manuals; descriptions of psychological journals; a catalogue of audiovisual materials; a listing of psychological equipment manufacturers; bibliographies of collateral readings for students and reference works for teachers; names and addresses of national organizations; a discussion of teaching methods; and suggestions of ways to organize a first course in psychology. The book is one of several products of a project held at Oberlin College in the summer of 1970. Although the title of the book – *PROGRAM ON THE TEACHING OF PSYCHOLOGY IN THE SECONDARY SCHOOL:* Final Report indicates an applicability to a particular academic level, the book is useful in teaching general psychology regardless of the level.

**Pages: 128**

**Price: $2.00**

..........................................................................................

SOURCE BOOK ORDER FORM

Name ..........................................................................................

Address ......................................................................................

..........................................................................................

..........................................................................ZIP CODE.............

Send this form to:

American Psychological Association
Order Department
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

Orders must be prepaid. Single copy $2.00. Quantity discounts: 3-10 copies 10%, 11-Over copies 20%. Make checks payable to the American Psychological Association.

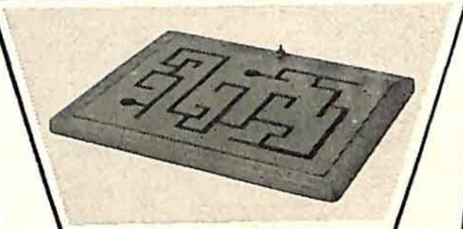
*Lafayette* **INSTRUMENT COMPANY**

PERCEPTUAL MOTOR

LEARNING

PHYSIOLOGICAL

**31010—Mirror Tracer**
Unbreakable Stainless Steel Mirror • Suitable for Either Right or Left-Handed Use • Completely Collapsible
$16.50



**20014—Pencil Maze**
Durable Aluminum Construction • Uses Standard 8½ x 11" Paper Provides a Permanent Record
$16.50

**77025—Student Galvanometer** • Completely Portable—Battery Operated Large Scale Ideal for Use with an Overhead Projector • Highly Sensitive
$119.00

*pointing the way in*
LABORATORY EQUIPMENT

Lafayette Instrument Company manufactures a full line of General Laboratory Equipment in all areas of concentration from basic devices designed for student operation to highly sophisticated instruments for research use. This includes a full line of Memory Drums, Learning Devices, Perceptual Motor Skill Devices, Manual Dexterity Tests, Sensory and Perceptual Equipment, Flicker-Fusion Units, Polygraphs, Reaction Time Devices, Tachistoscopes, Timers, Counters, Animal and Human Behavioral Control Instruments, and Programming Equipment. Lafayette Instrument Company also employs a complete staff of engineers, technicians, sales, and psychological personnel trained to aid you in your instrumentation needs in the most efficient, economical, and reliable means. Please feel free to contact us at any time.

**Lafayette Instrument Company**
P. O. BOX 1279 • 52 BY-PASS • LAFAYETTE, INDIANA 47902 • 317-423-1505

**JESSE JONES VOLUME FILES**

Send orders to:
JESSE JONES CORP.
P.O. Box 5120
Philadelphia, Pa. 19141

Only $3.50 ea.
3 for $10.00
6 for $19.00 P.P.D.

Now, the journals of the American Psychological Association can become an attractive permanent part of your professional library. These famous Jesse Jones volume files, especially designed to keep your copies orderly, readily accessible for future references — guard against soiling, tearing, wear or misplacement of copies.

This durable file will support 150 lbs. Looks and feels like leather and is washable. The 23-carat gold lettering makes it a fit companion for the most costly binding.

Journal of
Experimental Psychology

Vol. 93, No. 2 May 1972

# HUMAN PERFORMANCE IN CONTINGENT INFORMATION-PROCESSING TASKS [1]

IRVING BIEDERMAN [2]

*State University of New York at Buffalo*

Theoretical accounts of complex human information-processing behavior have emphasized the utilization of contingencies whereby the processing of some information directs the processing requirements of the remaining sources of uncertainty. The present investigation sought to determine how such flexible processing might be accomplished in a speeded recognition task where the relevancy of a given stimulus dimension was contingent upon the value of the stimulus on some other dimension. These contingent tasks, in which only two of the three dimensions were relevant on any one trial, were performed faster and more accurately than tasks where all three dimensions were relevant on each trial. Moreover, the speed at which a given contingent task could be performed was related to the discriminability of the dimensions that were relevant on that trial. These results provide strong support for a self-terminating, feature-testing, contingent mode of processing. The error data and repetition effects also supported the hypothesis that *S*s were employing the contingent relations to achieve a classification of the stimulus. The lack of interactions between discriminability of the relevant dimensions in the contingent tasks and variations in S–R compatibility were consistent with the hypothesis that the contingencies were utilized at some stage prior to response selection.

The experimental study of choice reaction time (RT) and perceptual recognition has typically involved performance on tasks where a given source of information was either always relevant or never relevant (Posner, 1964; Smith, 1968; Treisman, 1969). Outside of the laboratory, however, information which is relevant in one situation can often be ignored in another; i.e., most information that we deal with has the status of being *sometimes* relevant. Obviously, the relevancy or irrelevancy of this

---

[1] This research was submitted in partial fulfillment of the requirements for the PhD degree at the University of Michigan. The research was initiated under the guidance of the late Paul M. Fitts. Appreciation is expressed to the members of the dissertation committee: Arthur W. Melton, chairman, Wilfred M. Kincaid, Richard W. Pew, and Irwin Pollack for their kind encouragement and many helpful suggestions. Edward E. Smith and Gregory R. Lockhead provided careful readings of the manuscript. The research was supported by the Advanced Research Projects Agency, United States Department of Defense, and was monitored by the Air Force Office of Scientific Research under Contract AF 49(628) 1235 with the Human Performance Center, Department of Psychology, University of Michigan. The author held a United States Public Health Service predoctoral research fellowship during this period.

[2] Requests for reprints should be sent to Irving Biederman, Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Buffalo, New York 14226.

© 1972 by the American Psychological Association, Inc.

information must be *contingent* upon the status of other sources of information.

The processing of contingent information is of central importance in many theoretical accounts of complex information processing (Hunt, 1962; Miller, Galanter, & Pribram, 1960; Reitman, 1965; Trabasso, Rollins, & Shaughnessy, 1971). Contingent structures are often presented diagrammatically as tree diagrams or branching flow charts or verbally referenced as "hierarchical structures." Presumably, the functional significance of contingent (or hierarchical) structures resides in the economy which they afford in the required number or complexity of the information transformations intervening between input and output. In some sense, virtually all information-processing tasks have contingencies embedded in them in that if sources of stimulus information are sampled then an efficient system does not reexamine the already examined sources. In other words, the search among the remaining alternatives is contingent on the fact that the "critical" alternative has not, as yet, been identified. While the underlying psychological processes may be the same in the two cases, the term contingent will be used only with reference to those tasks where the processing of one source of information conveys more information about the necessity of processing remaining sources than would arise from simple elimination of that source as critical.

The present research is designed to examine the following three questions concerning the processing of contingent information. (*a*) Can contingent information be used to direct the course of multidimensional perceptual recognition? While it is obvious that contingent information is employed in many activities typically classified as thinking, it is not clear whether such information would be employed in the immediate classification of stimuli. (*b*) If contingent information is used, how is it handled? (*c*) At what stage in the information-processing sequence between stimulus and overt response does this processing occur?

*Experimental tasks.*—Discussion of the above three questions will be facilitated if reference is first made to a sample of the tasks employed in the experiment. The tasks used the same set of eight stimuli which were assigned to four finger-key responses: two stimuli per response (Fig. 1

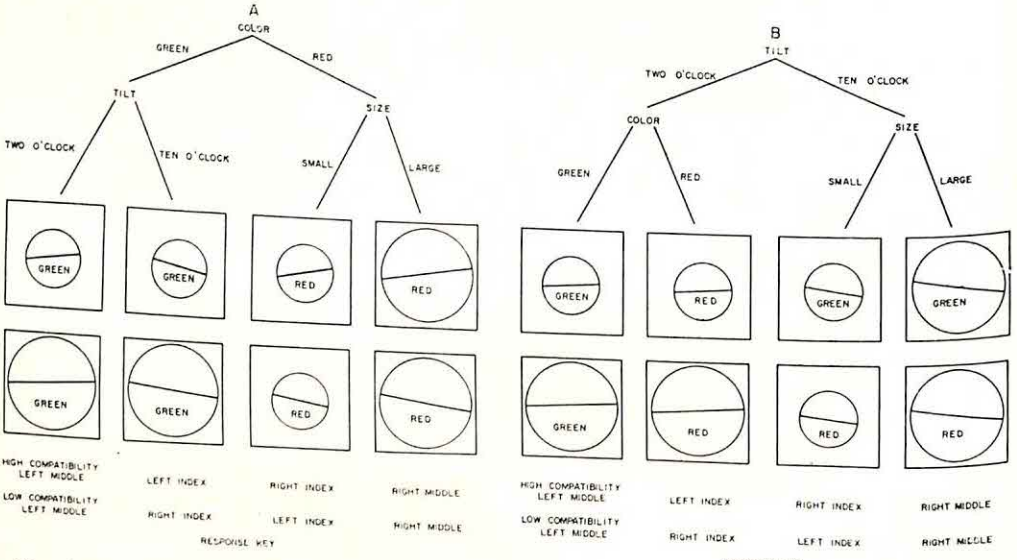


FIG. 1a. Schematic of a contingent task with color as a primary dimension (CC Set A) showing the S–R assignments for both high- and low-compatibility conditions. (1b. Schematic of a contingent task with tilt as a primary dimension (CT Set A). A serial contingent processing tree is shown above each task.)

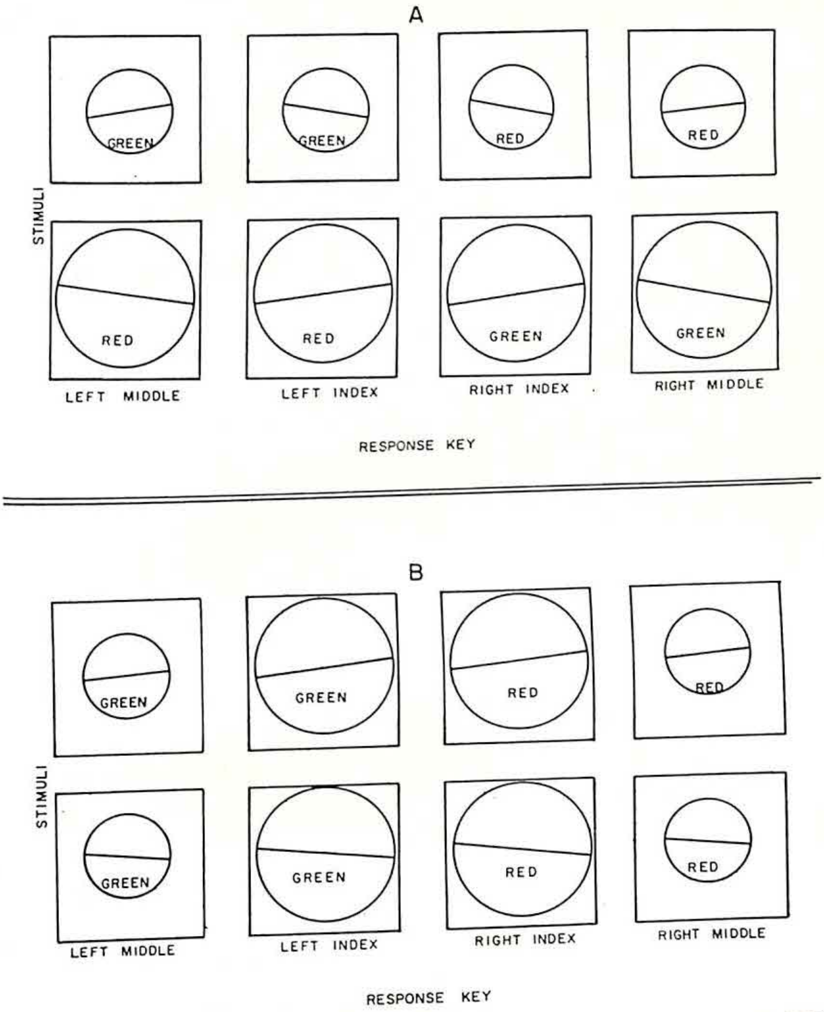


FIG. 2a. Schematic of the S–R assignments for a condensation task (COND Set A). (All three dimensions are relevant. 2b. Schematic of the S–R assignment for a filtering task (FT Set A) with color and size relevant and tilt irrelevant.)

and 2). The eight stimuli were generated from three binary dimensions (background color, circle size, and diameter tilt). The tasks differed only in the way these eight stimuli were assigned to the four responses. In the *contingent* tasks, the assignment was such that the relevance of a given *secondary* dimension was dependent upon (or contingent upon) the value of the stimulus on a *primary* stimulus dimension. Figure 1a shows a contingent task where color is the primary dimension and tilt and size are secondary dimensions. If the color is green, *S* can respond on the basis of tilt and ignore size. If the color is red, then *S* can respond on the basis of size and ignore tilt. Figure 1b shows a contingent task with the same set of stimuli and responses as in Fig. 1a, but with tilt as the primary dimension. In the *condensation* task (Fig. 2a), all three dimensions were always relevant. In the *filtering* task (Fig. 2b), one of the three is always irrelevant; the other two are always relevant.

*Contingent information processing.*—The three tasks shown in Fig. 1 and 2 are what Posner (1964) termed information-reduction tasks in that the response information is less than the stimulus information (although Posner distinguished only between filtering, which he called gating, and condensation). When relevant and irrelevant visual information are perceptually distinct or analyzable, a filtering task requires less processing time than a corresponding condensation task (Fitts & Biederman, 1965). Posner maintained that an extra transformation step is required to effect the reduction of information in the condensation task. Such a step would not be required in the filtering task where the information to be reduced is never relevant. In Posner's (1966) terms, then, the difficulty of the contingent task relative to a filtering or condensation task would depend on whether a transformation is performed on a secondary dimension on those trials when it is irrelevant.

Two studies have been explicitly concerned with RT in a contingent information-processing task (Hodge, 1959; Montague, 1965). In the Montague experiment, irrelevant information that was at other times relevant led to more errors in a contingent task than did the same amount of irrelevant (and never relevant) information in a filtering task. Montague interpreted this as indicating that "the locus of interference attributable to irrelevant information is in the response competition generated by implicit responses elicited by the nonrelevant dimensions [p. 235]." Hodge also accepted response competition as an explanation for the increase in reaction times (RTs) that occurred when the number of irrelevant, but at other times relevant, dimensions of a stimulus was increased. However, current models of information-processing behavior (e.g., Sternberg, 1969) hold that "implicit responses" occur at functionally different processing stages between input and output. It would seem desirable, therefore, to specify a more precise locus in the sequence of processing stages for the effects of sometimes relevant information. Also, as will be seen below, vastly different information-processing modes are consistent with the results of both the Hodge and Montague experiments.

There are at least two ways in which contingent relations among dimensions might be used to influence the course of perceptual identification. The different modes depend on whether dimensions are processed sequentially or in parallel (Biederman & Checkosky, 1970; Egeth, 1966; Nickerson, 1970). A sequential mode is illustrated by the decision tree in the upper portion of Fig. 1: *S* might use the value of the primary dimension as the basis for selecting the relevant secondary dimension for processing. To do this, it is necessary to process the primary dimension first, select one secondary dimension for processing, and ignore the other secondary dimension. This mode implies that the primary dimension would be processed on every trial while any one secondary dimension would be processed on only those trials in which it was relevant. Such a mode of processing will be termed *serial contingent processing*. The utility of employing serial contingent processing arises from the savings in time when one of the secondary dimensions is not processed.

If dimensions are processed in parallel, then the contingency relations might be used to define criteria for determining when a sufficient amount of information had been extracted from the secondary dimensions for a correct response. This mode, which will be termed *parallel contingent processing*, would benefit *S* only on those trials when the processing of the primary dimension was completed before the completion of the irrelevant secondary dimension. For example, in the task shown in Fig. 1a, *S* would simultaneously start to process color, size, and tilt. If *S* determined that the color of the stimulus is red, say, then the parallel contingent mode implies that *S* has the capacity to initiate a successive stage of processing (e.g., response selection) as soon as he determined the size. Even though the tilt is not "known" to be irrelevant, *S* may continue to process it. However, the outcome of the tilt processing, even if it should come prior to the outcome of the size processing, would not influence the initiation of a successive stage.

Note that both contingent modes assume a highly flexible process in which early outcomes of processing are rapidly and efficiently utilized to influence the subsequent course of processing. Biederman and Checkosky (1970) argued for such flexibility in showing that when stimuli differed on two dimensions, either of which could furnish sufficient information for a correct response, RTs were faster than when stimuli differed on only one dimension. A model compatible with this result is one in which the different dimensions are processed in parallel with a response initiated as soon as sufficient information is available. Given some variability, overlap, and independence in the distribution of times for processing each dimension, sometimes the processing for one and sometimes the processing for the other dimension would be completed first. To capitalize on this variability, *S* could not adopt a stopping rule based on a dimension specified prior to processing. Rather, he would have to start processing both dimensions and to initiate response selection based on whatever dimension was completed fiirst. Similarly, Neisser (1967) has argued that the high speed and accuracy evidenced when one identifies a stimulus requires that a preliminary analysis be made of certain features of the stimulus. This preliminary analysis serves to guide the course of subsequent feature testing.

Selecting secondary dimensions to process or to use as criteria for stopping rules is an activity that in itself, however, might consume more time than is saved by not processing an irrelevant dimension. There are a number of ways in which *S* might *not* use the contingency relations and still correctly and efficiently perform the task. Four such modes will be considered.

1. The dimensions might be processed sequentially, in the order of primary, easy secondary, and difficult secondary (assuming that the secondary dimensions differed in processing difficulty). While this mode, which will be termed *sequential by primary*, does capitalize on the contingent relations formed by the S–R assignments, it is not termed a contingent mode since the outcome of prior processing does not serve to influence the course of subsequent processing.

2. The dimensions might be processed sequentially, in order of their discrimination difficulty. With this mode, which will be termed *sequential by discriminability*, not only is the relevancy or irrelevancy status of the secondary dimensions ignored, as with the sequential-by-primary mode, but the primary versus secondary status of the dimensions is also ignored.

3. All three dimensions might be processed on every trial *exhaustively* even when sufficient information is available from two of the dimensions to make a correct response. Both contingent modes, as well as the sequential-by-primary and sequential-by-discriminability modes, assume that a successive stage can be initiated as soon as sufficient information is available. For purposes of the present experiment, it is unnecessary to distinguish between parallel or sequential forms of exhaustive processing.

4. While the set of stimuli may be represented as being generated from several dimensions, it is not necessarily the case that *S* "analyzes" a stimulus into its component dimensions. Thus, *S* might form some internal, unitary "template" of the stimulus (visual, verbal, or abstract) in which individual features are not tested but in which the complete stimulus is the unit for input–output matching. Since the contingencies being discussed require testing features of the input, contingent processing could not be manifested if the internal representation of the stimulus was in template form.

The experimental strategy to distinguish among the various processing modes was to compare RTs from different contingent tasks where differences in discriminability (physical similarity) were interchanged between primary and secondary dimensions. To accomplish this, while using the same set of stimuli in the different tasks, the stimulus set was generated from dimensions that differed in the discriminability of their values and the contingent tasks differed in the dimension which was primary. With the tasks shown in Fig. 1 and 2, the tilt dimension was harder to discriminate than the color or size dimensions. The expected relations between RTs and primary dimension discriminability discussed below are based on the assumption that differences in RT between tasks reflect the number of dimensions that are processed and the difficulty in making discriminations on these dimensions.

Both exhaustive processing and template matching imply that there would be no difference in RTs between contingent tasks and condensation tasks. In both kinds of tasks, all three dimensions (or all eight templates) would be processed on every trial. The sequential-by-discriminability mode in which the dimensions were processed in order of easiest to most difficult implies that there would be no difference between a condensation task and a contingent task in which the most difficult dimension was primary (and last to be processed). In both tasks, RTs would be limited by the same third dimension to be processed. If the dimensions were processed in order of most difficult to easiest, then no difference would obtain between a contingent task with the least difficult dimension as primary and a condensation task. Moreover, this version of the sequential-by-discriminability mode predicts that the more difficult the discrimination of the values of the primary dimension, the *faster* the RTs.

The sequential-by-primary mode implies that if a contingent task had the most difficult dimension as a secondary dimension, RTs on those trials when that difficult dimension was relevant would be equal to RTs on a condensation task. For example, with the task shown in Fig. 1a, when a green stimulus is presented, *S* would first process its color, then its size (even when irrelevant), and finally its tilt (the most difficult dimension). Thus all three dimensions would be processed as is required in the condensation task. If the differences in discriminability are so great that there is no overlap between the component distribution of times for identifying the values on the dimension that was most difficult to discriminate and the distribution of component times of the dimension that was easier to discriminate, then the parallel contingent mode would also yield the same prediction as the sequential-by-primary mode.

The serial and parallel contingent modes can also be distinguished by means of an analysis of errors. Many models of information-processing behavior view errors as resulting from a premature termination of a sequential search or sampling process (Egeth & Smith, 1967; Fitts, 1966). Under the serial contingent mode, since the primary dimension is processed first, there would be fewer errors attributable to misidentifications of the primary dimension (*primary errors*) than to misidentifications of the secondary dimensions (*secondary errors*). Such an outcome could also be consistent with the parallel contingent mode if it is assumed that a response would tend to be initiated once the value of the primary dimension was determined. Given that RTs are related to the discriminability of the primary dimension and that there are more primary errors than secondary errors, the parallel contingent model predicts that the secondary error rate would be related to the discriminability of the primary dimension. If *S* starts out by processing all three dimensions simultaneously, and if the primary dimension required relatively little time to discriminate, then the processing of the secondary dimensions would tend not to be completed until after the processing of the primary dimension was completed. Now, if *S* initiated, on some trials, a response as soon as he determined the value of the primary dimension, then he would be likely to misidentify the value of the secondary dimension. Alternatively, if the primary dimension was difficult to discriminate, then by the time *S* completed the processing of the primary dimension he would be likely to have already completed the processing of the secondary dimensions. Consequently, in this latter case he would not be likely to misidentify the value of the secondary dimension. Thus the parallel contingent model predicts that there would be more secondary errors on a given dimension when that dimension is in a task with a primary dimension that is relatively easy to discriminate than when it is in a task with a primary dimension that is relatively difficult to discriminate. For example, in the tasks shown in Fig. 1, the parallel contingent model predicts that more errors would be made in misidentifying size when it was a secondary dimension in the CC task (where a relatively easy dimension, color, is primary) than in the CT task (where a difficult dimension, tilt, is primary).

*Repetition effects.*—The existence of repetition effects (e.g., Bertelson, 1963) has provided investigators of RT with another dimension to their dependent variable. Schvaneveldt and Chase (1969) have recently argued that these effects reflect processing strategies. If so, then of particular interest in the study of contingent tasks is the possibility that the magnitude of the repetition effect may be sensitive to the informational (or functional) significance of the different stimulus components. Shaffer (1965) reported a study in which he argued that repetition effects were more marked with stimuli designated as rules than with stimuli designated as signals to which rules were to be applied. The signals were two lights arranged horizontally to correspond to the spatial separation of the two index fingers. The "rule" was comprised of two intersecting bars at right angles to each other to form a "+." When the horizontal bar was presented, a homolateral S–R mapping of signals to index fingers was defined; i.e., the left light was to be responded to with the left index finger and the right light was to be responded to with the right index finger. When the vertical bar was presented, a contralateral S–R mapping was defined.

In Shaffer's (1965) study, RTs were faster when either the rule or the signal was repeated on successive trials, but repeating the rule led to a greater facilitative effect than repeating the signal. There are some difficulties, however, in generalizing Shaffer's results. If the utility attendant upon the use of rules and concepts is to be understood in terms of their provision for efficient processing of sets of alternatives (whether these alternatives be additional features of the stimulus situation or additional S–R mappings), then Shaffer's designation of the bars as rules and the lights as signals would appear to be arbitrary: The bar conveyed as much information as the light, independently of the order in which the two stimulus elements were processed. Another difficulty in interpreting Shaffer's results is that the differences in physical characteristics of the two types of stimuli were confounded with their rule versus signal designations. A more adequate study of the use of rules in information-processing behavior would involve use of a task in which initial extraction of information from one stimulus dimension (designated as the primary or rule dimension) results in less total information processing than if some other (secondary) dimension is processed first. The contingent tasks shown in Fig. 1 meet this specification. Moreover, functional effects may be distinguished from the effects of the different physical characteristics of the stimulus dimensions by counterbalancing the dimensions assigned as rules and signals.

Shaffer's (1965) question as to the relative magnitude of the repetition effect for rules versus signals to which the rules are to be applied may be profitably asked about the performance of contingent tasks. Under the serial contingent mode, it is clear that the primary dimension would be classified as the rule and the secondary dimensions as the signals. Thus, repeating the value of the primary dimension might result in a repetition effect of greater magnitude than would be obtained if only the secondary dimension was repeated.

*Locus of effects in the information-processing sequence.*—In the tasks studied, two response dimensions can be identified: hand (left vs. right) and finger (index vs. middle), with the former generally leading to more potent compatibility manipulations. S–R compatibility of the contingent tasks was varied by utilizing an S–R mapping in which the value of the primary dimensions determined not only which of the two secondary dimensions was relevant but also which hand was to be used in the high-compatibility conditions. In the low-compatibility conditions, the value of the primary dimension determined neither the hand nor the finger that was to be used. These variations are shown in the lower portion of Fig. 1a and 1b. If the contingent mode of processing primarily entails a selection of response, instead of stimulus, dimensions, then the utilization of an S–R coding in which response selection is not directly attainable from an identification of the primary dimension should reduce the facilitative effect of contingent processing. That is, there would be less evidence for a contingent mode of processing under the low-compatibility conditions. If the contingencies are utilized at some stage prior to response selection, then the effects which provide evidence for contingent processing should not interact with compatibility (Biederman & Kaplan, 1970).

By using an information-reduction task in which more than one stimulus is assigned to a single response, there are Trial $N$ to Trial $N + 1$ sequences in which different stimuli occur but in which the same response is correctly given to both. RTs from these sequences can be compared to those from sequences in which the identical stimulus is repeated on the two trials. If RTs from the latter are faster, it can be taken as evidence that processes prior to response selection are involved in the production of the repetition effect.

## METHOD

*Stimuli.*—The stimuli were circles containing a diameter on a colored square background. Eight stimuli were generated from the three binary dimensions of circle size, background color, and the angular orientation of the diameter (Fig. 1). Two sets of stimuli were used so that variations in discriminability (via physical similarity) could be isolated from effects associated with the specific dimensions. The values of the size dimensions were the same in both sets, the circles were either 36.5 mm. or 29.0 mm. in diameter. The two sets differed only with respect to the discriminability of the values of the background color and diameter tilt dimensions. In one set (Set A), the colors were either green or red (Kodak Wratten Filters No. 55 and 25, respectively), and the diameters were tilted ±2.25° from the horizontal so the angular difference between the two values was only 4.5°. If the right part of the diameter was above the horizontal, it was referred to as "two o'clock." If the left side of the diameter was above the horizontal, it was referred to as "ten o'clock." As measured by a MacBeth Illuminometer, the background colors had a luminance of approximately 3.0 ftc. In Stimulus Set B, the colors were slightly different shades of red, with one red identical to the one used in Set A. The other red had a dominant wavelength closer to orange (Kodak Wratten Filter No. 23A) and slightly more luminance (3.5 ftc.). This color was referred to as "orange." In Set B, the diameters were tilted ±45° from the horizontal. The angular difference between the two values in Set B was 90° and highly discriminable.

The circles and the diameters appeared as white outlines on the colored square; the color was therefore visible in the interior of the circle. The colored background squares were 50.8 mm. on a side. Stimuli were presented by a multiunit projection apparatus which permitted the independent display of two slides on each trial (Mylrea, 1966): one for background color and the other for a particular circle size-diameter tilt combination. The images from the two slides were superimposed on a ground-glass screen approximately 71 cm. from *S*'s eyes. The circles were centered within the background square.

To ascertain the relative discriminability of the different dimensions, six *S*s performed two choice RT tasks in which one of the dimensions was relevant and the other two dimensions were held constant. The stimuli from Set A were used with three *S*s; the other three *S*s had Set B stimuli. In both cases, mean RTs were fastest when size was the relevant dimension. In Set A, color was faster than tilt. In Set B, tilt was faster than color.

*Responses.*—A response consisted of the depression of one of four piano-like microswitch keys arranged so as to correspond to the natural placement of the middle and index fingers of the left and right hands (see Fitts, 1966, Fig. 2 for a diagram of the response panel).

*Procedure.*—The procedure insured that *S* learned the assignment of stimuli to responses. A card with a facsimile of the eight patterns, labeled and arranged according to the correct S–R pairings, remained directly in front of *S*'s response panel throughout an 8-min. instruction period and the first block of 128 trials. The *S* was told the rules for performing the contingencies, e.g., "if the pattern is green, then you need only consider the diameter tilt to determine which key is correct," or for performing the filtering, e.g., "only the size and the color need be considered to perform correctly; the diameter tilt may be ignored." A 3-min. period was allotted for *S* to study the card after these instructions were given. After this study period, the eight stimuli were projected and the values on the three dimensions were identified verbally by *E*.

An experimental session consisted of five blocks of 128 trials. The first block was termed a "practice block," and *S* was encouraged to strive for accuracy while committing the task to memory. A verbal ready signal initiated each block. The task was self-paced; patterns remained displayed until *S* made a response. Simultaneous with the offset of the pattern, a 6.4-mm. fixation cross would appear at the position occupied by the center of the circle. The interval after a response and before the next stimulus was 2 sec. During this time, one of six feedback lights would flash indicating whether the response was correct or incorrect and whether it was fast, intermediate, or slow. The fast and inter-

TABLE 1

S–R ASSIGNMENTS OF THE 10 TASKS OF STIMULUS SET A

| Stimuli | Contingent task | | | | | | Filtering task | | | Condensation task |
|---|---|---|---|---|---|---|---|---|---|---|
| | CSH | CSL | CCH | CCL | CTH | CTL | FT | FC | FS | Cond. |
| S G 2 | LM | LM | LM | LM | LM | LM | LM | LM | LM | LM |
| S G 10 | LM | LM | LI | RI | RI | LI | LM | RM | LI | LI |
| S R 2 | LI | RI | RI | LI | LI | RI | RM | LM | RM | RM |
| S R 10 | LI | RI | RI | LI | RI | LI | RM | RM | RI | RI |
| L G 2 | RI | LI | LM | LM | LM | LM | LI | LI | LM | RI |
| L G 10 | RM | RM | LI | RI | RM | RM | LI | RI | LI | RM |
| L R 2 | RI | LI | RM | RM | LI | RI | RI | LI | RM | LI |
| L R 10 | RM | RM | RM | RM | RM | RM | RI | RI | RI | LM |

Note.—Stimulus coding: S = small, L = large, G = green, R = red, 2 = two o'clock, 10 = ten o'clock. Response coding: L = left hand, R = right hand, M = middle finger, I = index finger.

mediate criteria were 375 and 525 msec. for the filtering and contingent tasks and 600 and 800 msec. for the condensation tasks. A pilot study revealed that the speed criteria specified for the contingent and filtering tasks would be unattainable in the condensation tasks. After each block, there was a 2-min. rest interval during which *S* was given a score sheet indicating the number of times he received each of the six feedback lights on the preceding block.

The *S*s were instructed to strive for accuracy rather than speed on the practice block but after that to "go fast enough so that about 5 per cent of your responses are in error." The *S* was reminded of this if his error rate exceeded 25%.

To reduce visual and auditory distraction, *S* sat in a dimly lit room and wore earphones into which white noise was fed. The earphones were part of an intercom system which enabled *E* and *S* to communicate from their separate rooms.

Stimulus sequences were presented by a punch-tape system. A different sequence was used for each block. The sequences were generated so that each of the eight stimuli appeared eight times in both halves of each block. The 64 possible Trial *N* to Trial $N + 1$ stimulus pairs yielded approximately equal frequencies of the 16 possible Trial *N* to Trial $N + 1$ response pairs.

*Tasks.*—There were 10 tasks within each stimulus set: 6 of the 10 tasks were contingent tasks, 3 were filtering tasks and 1 was a condensation task. Table 1 shows the S–R assignments for the 10 tasks of Set A. Contingent tasks are symbolized by an initial "C" followed by the letter indicating the dimension which is primary: "C" for color, "S" for size, and "T" for tilt. The last letter, "H" or "L," indicates whether the task was high or low, respectively, in compatibility. Thus "CSL" indicates a contingent task with the size dimension as primary and which was low in compatibility. Filtering tasks are symbolized by an initial "F" followed by the letter indicating the dimension which was to be ignored. Thus "FC" refers to a filtering task in which color was the irrelevant dimension (tilt and size, therefore, would be the relevant dimensions). The condensation task is symbolized by "COND." Table 1 also shows the S–R assignments of the 10 tasks of Set B with the one modification that "O" for orange be substituted for "G" for green. (While the differences between the diameter tilts were not the same in the two sets, the "two o'clock" and "ten o'clock" labels need not be changed since they correctly designate the quadrants of the upper portion of the diameters for both sets.)

Eight stimuli were assigned to four responses in all 10 tasks represented in Table 1. The low-compatibility contingent tasks were derived from their high-compatibility counterparts by switching the stimuli assigned to the index fingers.

*Design.*—The *S*s were assigned to 1 of the 10 tasks within a set. The 6 contingent tasks comprise a mixed design with two between-group variables and three within-*S* variables. The two between-group variables were compatibility (high vs. low) and primary dimension (color vs. size vs. tilt). One of the three within-*S* variables was the relative discriminability (high vs. low) of the relevant secondary dimension. The other two within-*S* variables were those for repetition effects. One of these variables was whether the value of the primary dimension changed from Trial *N* to Trial $N + 1$, the other was whether none, one, or both secondary dimensions changed value from Trial *N* to Trial $N + 1$.

The data of the filtering tasks were analyzed in a manner analogous to the contingent task analysis. In this analysis, it was assumed that the "primary" dimension was the one that determined which hand would be used, and the "secondary" dimension was the one that determined which finger would be used. A hand variable (left vs. right) was substituted in the filtering task analysis for the within-task discriminability variable in the contingent task analysis. A change from Trial *N* to Trial $N + 1$ in the value of the "primary" dimension in the filtering tasks, as in the high-compatibility contingent tasks, required a change in the hand making the correct response. Since in the high-compatibility contingent tasks the secondary dimension which was high in discriminability was assigned to the right hand in five out of six tasks, both the filtering tasks and the low-compatibility contingent tasks provided a control for hand effects.

*Subjects.*—The *S*s were 188 University of Michigan undergraduate and graduate students who volunteered for paid participation. Ten *S*s were assigned to each of the three filtering and six contingent tasks in Set A and in Set B. Four *S*s were assigned to each condensation task in each set. An equal number of males and females was assigned to each task.

*Data analysis.*—For each *S* in the contingent tasks, an 8 × 8 matrix of mean RTs of correct responses on Trial $N + 1$ to all 64 possible stimulus shifts from Trial $N$ to Trial $N + 1$ was constructed from the last three blocks of 128 trials. The first RT from each block was discarded because it had no immediate antecedent. A mean of the four means of the RTs to a particular value of the primary dimension on Trial $N + 1$ was computed. Thus the 8 × 8 matrix was collapsed into an 8 × 2 data matrix by ignoring differences in RTs between the four stimuli having the same value on the primary dimension on Trial $N + 1$. The 16 means were the data points used, after log transformation, in the analysis of variance. Each of the 16 means represented one combination of treatment levels in a 2 × 2 × 4 within-*S* factorial design in which the first variable was the relative discriminability of the secondary dimension that was relevant on Trial $N + 1$; the second variable was whether the primary dimension changed from Trial $N$ to Trial $N + 1$;

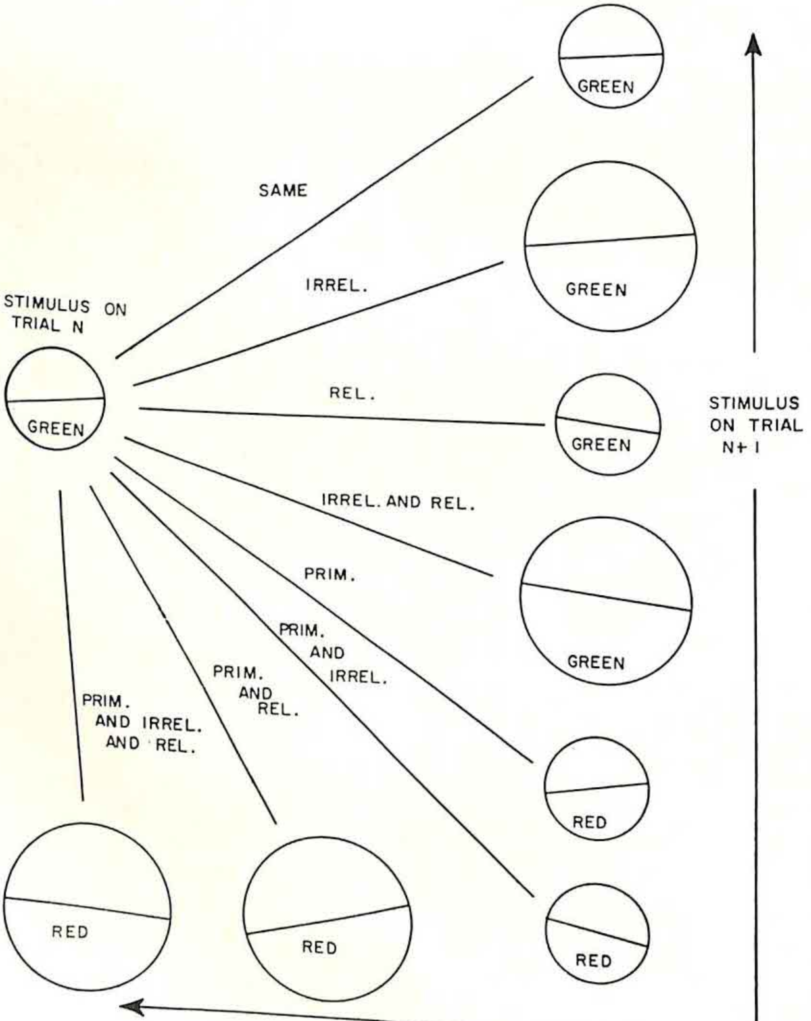


FIG. 3. Trial $N$ to $N + 1$ stimulus sequences showing categorization of stimulus changes for analysis of repetition effects. (Example is from Contingent Task CC (Set A) with color as primary dimension. Irrel. and rel. refer to changes in value(s) of secondary dimensions with respect to their status on Trial $N + 1$. Prim. refers to changes in value of primary dimension.)

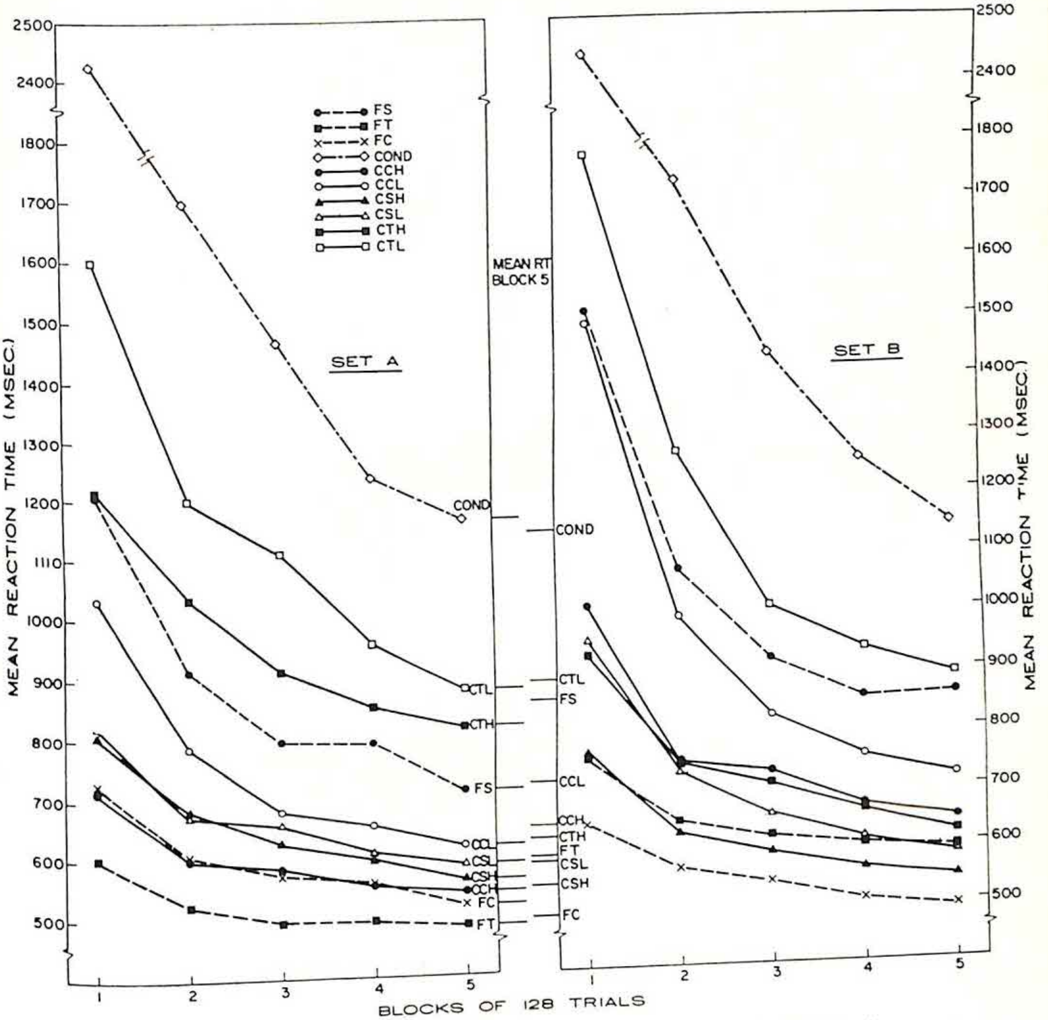


Fig. 4. Mean correct RT as a function of task and trial block.

and the third variable was the change from Trial $N$ to Trial $N + 1$ in the values of the secondary dimensions. There were four levels of this last variable: (*a*) both secondary dimensions had the same value on Trial $N + 1$ as they had on Trial $N$, (*b*) the irrelevant secondary dimension on Trial $N + 1$ changed its value from Trial $N$, but the relevant secondary dimension did not change value, (*c*) the relevant secondary dimension on Trial $N + 1$ changed its value from Trial $N$ but the irrelevant secondary dimension did not change value, and (*d*) both secondary dimensions changed value from the preceding trial. These categories are illustrated in Fig. 3.

## Results

*Tasks.*—Figure 4 shows that the task differences were relatively stable over practice. The condensation tasks were clearly the slowest in both sets (and also had the highest error rates). In general the contingent tasks were next in difficulty and the filtering tasks were easiest. The one exception to this is the FS task which was more difficult than most of the contingent tasks in both sets. Performance on this task was more comparable to the CTL task than to the filtering tasks.

*Discriminability.*—Figures 5 and 6 show that the orderings of the means of the contingent tasks were generally consistent with the orderings based on the discriminability of their primary dimensions. The differences among primary dimensions were highly significant, $F$ (2, 54) = 42.91, $p < .001$, and 20.19, $p < .001$, for Sets A and B, respectively. The exceptions to the orderings based on primary dimension

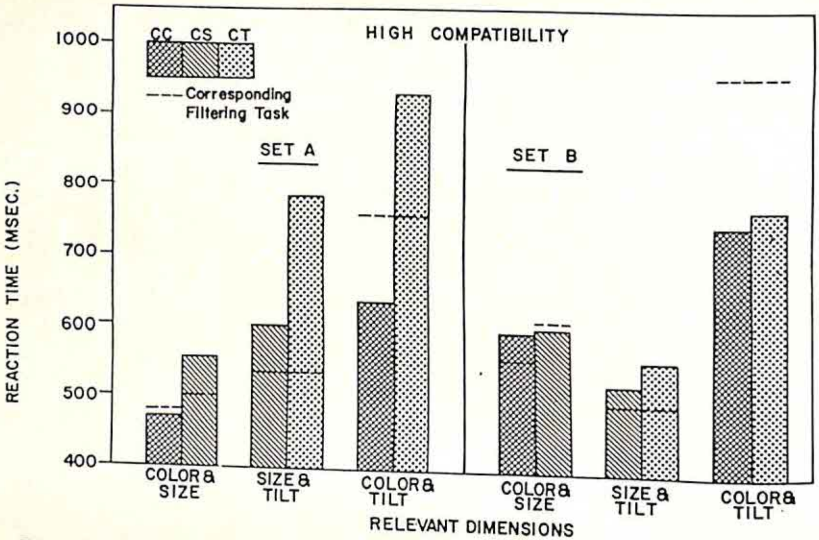


Fig. 5. Mean correct RTs as a function of pairs of relevant dimensions (one primary and one secondary) from the last three 128-trial blocks for the high-compatibility contingent tasks. (Code in upper left-hand corner specifies the contingent task of the pair: C = color primary, CS = size primary, CT = tilt primary. Dashed lines show corresponding RTs, matched by fingers, from filtering tasks with the same relevant dimensions.)

discriminability were that CCH was faster than CSH in Set A and that CCL was faster than CTL in Set B. Differences in RTs among the filtering tasks were also highly reliable, $F$ (2, 27) = 12.16, $p < .001$, and 23.94, $p < .001$, for Sets A and B, respectively. Within both sets, the ordering of RTs for the filtering tasks was consistent with the discriminability of the relevant dimensions.

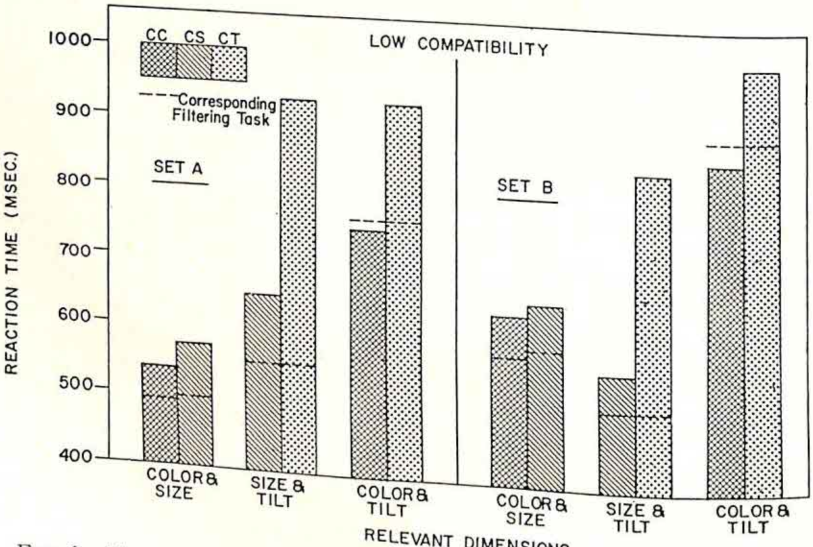


Fig. 6. Mean correct RTs as a function of pairs of relevant dimensions (one primary and one secondary) from the last three 128-trial blocks for the low-compatibility contingent tasks. (Code in upper left-hand corner specifies the contingent task of the pair: CC = color primary, CS = size primary, CT = tilt primary. Dashed lines show corresponding RTs, matched by fingers, from filtering tasks with the same relevant dimensions.)

Figures 5 and 6 show mean RTs grouped in terms of the pair of dimensions that were relevant on that trial. In the contingent tasks in Set A, RTs from those trials when the more discriminable secondary dimension was relevant averaged 95 msec. faster than RTs from those trials when the less discriminable secondary dimension was relevant, $F\ (1, 54) = 38.06, p < .001$. In Set B, this difference was 148 msec., $F\ (1, 54) = 97.89, p < .001$. It should be noted that when the least discriminable dimensions were relevant, i.e., tilt and color, RTs were still considerably faster than RTs on the condensation task (Fig. 4).

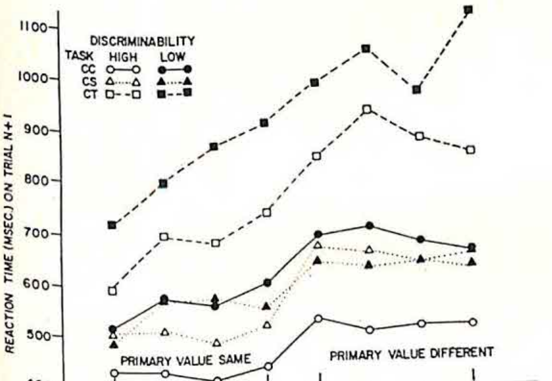


FIG. 7. Mean correct RTs from the last three 128-trial blocks for the contingent tasks from Set A. (The abscissa refers to the status on Trial $N + 1$ of the dimension(s) which changed value from Trial $N$ to Trial $N + 1$. [None = same secondary values on Trial $N$ and Trial $N + 1$.] $a$ = high-compatibility tasks. $b$ = low-compatibility tasks.)

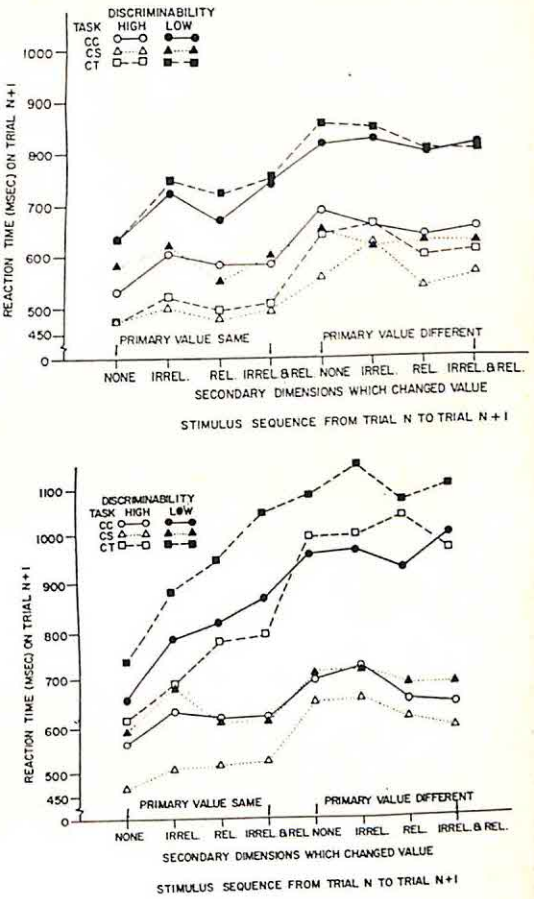


FIG. 8. Mean correct RTs from the last three 128-trial blocks for the contingent tasks from Set B. (The abscissa refers to the status on Trial $N + 1$ of the dimension(s) which changed value from Trial $N$ to Trial $N + 1$. [None = same secondary values on Trial $N$ and Trial $N + 1$.] a = high-compatibility tasks. b = low-compatibility tasks.)

*Compatibility*.—In all cases, the low-compatibility tasks had longer RTs than the corresponding high-compatibility tasks (Fig. 4). In Set A, this main effect of compatibility fell short of significance, $F\ (1, 54) = 3.91, .05 < p < .10$, but is significant in Set B, $F\ (1, 54) = 20.49, p < .001$. The relatively large difference of 261 msec. between the CTH and CTL tasks in Set B resulted in a significant Primary Dimension × Compatibility interaction, $F\ (2, 54) = 5.74, p < .001$. In both sets, the $F$ ratios for the Compatibility × Secondary Dimension Discriminability interaction were less than 1.00. This indicates that the above-mentioned main effect of secondary dimension discriminability in the high-compatibility tasks cannot be attributed to a left- versus right-hand effect.

*Repetition effects.*—The pattern of repetition effects was similar in the two sets as is shown in Fig. 7 and 8. In the contingent tasks, RTs averaged 130 msec. longer when the primary dimension changed value from the preceding trial than when its value was repeated, $F$ (1, 54) = 164.49, $p < .001$, in Set A, and 169.76, $p < .001$, in Set B. In the high-compatibility contingent tasks, such changes were confounded with hand changes. However, hand changes are insufficient to account for the effect since there was only a negligible increase in RTs when hand changes occurred in the FT and FC filtering tasks (Fig. 9 and 10). (The apparent exception is with FS which will be discussed later.) Moreover, the magnitude of the effect was approximately the same in the low- as in the high-compatibility contingent tasks; the Primary Value Change × Compatibility interactions were not significant, $F$ (1, 54) = 1.99 in Set A and 2.46 in Set B. In the low-compatibility tasks, hand changes were required on only half of the trials when the primary dimension changed value. RTs from these trials were indistinguishable from RTs when the primary dimension changed value and a hand change was not required.

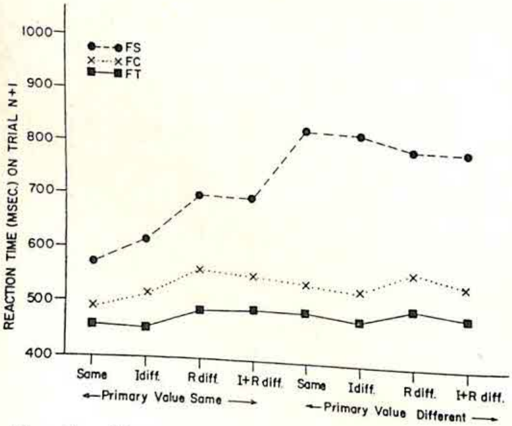


FIG. 9. Mean correct RTs from the last three 128-trial blocks for the filtering tasks from Set A. (The abscissa refers to the status on the $N$ + first trial of the dimension(s) which changed value from Trial $N$ to Trial $N + 1$. [Same = both secondary values repeated, I diff = irrelevant secondary value change, R diff = relevant secondary value change].)

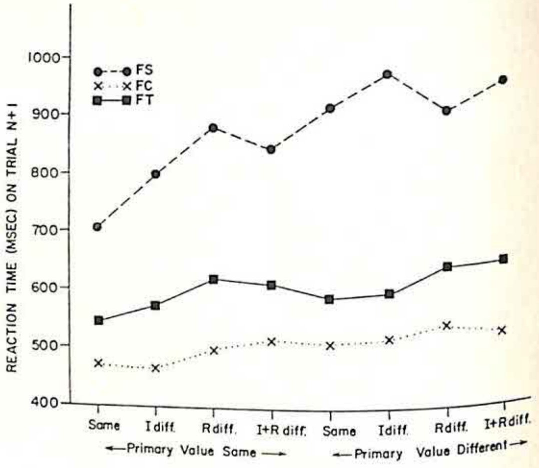


FIG. 10. Mean correct RTs from the last three 128-trial blocks for the filtering tasks from Set B. (The abscissa refers to the status on the $N$ + first trial of the dimension(s) which changed value from Trial $N$ to Trial $N + 1$. [Same = both secondary values repeated, I diff = irrelevant secondary value change, R diff = relevant secondary value change].)

When only an irrelevant dimension changed value from Trial $N$ to Trial $N + 1$, RTs were longer than when the stimulus was repeated on successive trials. In both cases, repetition of the response was required. The magnitude of this repetition effect was directly related to the discriminability of the irrelevant dimension that changed value and it occurred in both the contingent and the filtering tasks (Fig. 7, 8, 9, and 10).

*Errors.*—Each contingent task error was classified in terms of (*a*) whether the error could be attributed to an incorrect identification of the value of the primary dimension, of the secondary dimension, or of both, and (*b*) whether the secondary dimension which was relevant on the trial in which the error was made was high or low in discriminability. For example, in the high-compatible CC task in Set A (Fig. 1), if a small green two o'clock stimulus was presented and $S$ responded with the left index finger, then the error was classified as a secondary error because a misidentification of the value of the relevant secondary dimension would have resulted in that response. If *S* responded with his right index finger, then the error was classified as primary because if *S* has misidentified the primary dimension, he would have responded on the basis of size. Errors that involved misidentifications of both the primary and the irrelevant secondary dimensions (*S* would have made this type of error if he had responded with his right middle finger) constituted only 6.5% of the total number of errors and were therefore omitted from the following analysis.

Table 2 shows the mean error frequencies for the last three blocks of 128 trials for the two-way classification of primary dimension and compatibility for both sets of contingent tasks. Within both sets there were significantly more secondary than primary errors: $F\ (1, 54) = 51.90$, $p < .001$ for Set A, and $F\ (1, 54) = 27.81$, $p < .001$, for Set B. The error rates were also related to the discriminability of the different dimensions. Of the 1,738 primary and secondary errors in Set A, 409 involved misidentifications of size, 567 of color, and 762 of tilt. The comparable figures for the 1,934 errors of Set B were 342 for size, 1,106 for color, and 486 for tilt. These differences led to a significant secondary dimension discriminability effect, $F\ (1, 54) = 8.47$, $p < .001$, in Set A, and $F\ (1, 54) = 34.94$, $p < .001$, in Set B. Most important, there was only a negligible difference in the secondary error rates as a function of the discriminability of the primary dimension: an average of 9.8 errors could be attributed to misidentifications of secondary dimensions when in a task with a primary dimension that was relatively easy to discriminate and 10.5 errors when with a primary dimension that was relatively difficult to discriminate. More generally, the two types of error rates did not interact with dimension discriminability as shown in Table 3. Further categorization of the data by compatibility and stimulus set does little to alter this picture of independence of discriminability and error type.

TABLE 2

MEAN SECONDARY AND PRIMARY ERROR FREQUENCIES AS A FUNCTION OF PRIMARY DIMENSION, COMPATIBILITY, AND SECONDARY DIMENSION DISCRIMINABILITY FOR THE CONTINGENT TASKS OF HIGH AND LOW COMPATIBILITY IN SETS A AND B

| Task | Low discriminability | | High discriminability | | Total |
|---|---|---|---|---|---|
| | Secondary | Primary | Secondary | Primary | |
| Set A | | | | | |
| High compatibility | | | | | |
| CC | 15.0 | 3.6 | 9.9 | 1.6 | 30.1 |
| CS | 13.2 | 1.1 | 12.0 | 1.8 | 28.1 |
| CT | 12.4 | 6.7 | 7.7 | 3.3 | 30.1 |
| $\bar{X}$ | 13.5 | 3.8 | 9.9 | 2.2 | |
| Low compatibility | | | | | |
| CC | 12.9 | 6.0 | 4.5 | 2.6 | 26.0 |
| CS | 6.8 | 2.2 | 10.8 | 2.7 | 22.5 |
| CT | 7.7 | 9.9 | 11.0 | 8.4 | 37.0 |
| $\bar{X}$ | 9.2 | 6.0 | 8.7 | 4.1 | |
| Set B | | | | | |
| High compatibility | | | | | |
| CC | 9.2 | 15.8 | 6.8 | 7.6 | 39.4 |
| CS | 17.9 | 2.6 | 10.2 | 1.1 | 31.8 |
| CT | 22.2 | 2.2 | 7.9 | 1.2 | 33.5 |
| $\bar{X}$ | 16.4 | 6.9 | 8.3 | 3.3 | |
| Low compatibility | | | | | |
| CC | 6.7 | 9.1 | 3.2 | 9.6 | 28.6 |
| CS | 14.3 | 5.0 | 4.7 | 2.6 | 26.6 |
| CT | 14.1 | 6.8 | 5.0 | 7.6 | 33.5 |
| $\bar{X}$ | 11.7 | 7.0 | 4.3 | 6.6 | |

Note.—CC = color primary, CS = size primary, CT = tilt primary.

TABLE 3

MEAN PRIMARY AND SECONDARY ERROR FREQUENCIES AS A FUNCTION OF DIMENSION DISCRIMINABILITY

| Error type | Relative discriminability of misidentified dimension | | |
|---|---|---|---|
| | Easy | Intermediate | Difficult |
| Primary | 4.8 | 7.9 | 17.6 |
| Secondary | 16.5 | 18.6 | 29.1 |

Note.—Size was the easiest dimension to discriminate in both stimulus sets. In Set A, color was intermediate in discrimination difficulty and tilt was most difficult; in Set B, tilt was intermediate in discrimination difficulty and color was most difficult.

## DISCUSSION

*Modes of processing.*—The results strongly support a self-terminating, feature-testing, contingent mode of processing. The differences among contingent tasks differing in primary dimension discriminability, as well as between contingent and condensation tasks, support a self-terminating, feature-testing mode rather than a template matching or exhaustive process. The sequential-by-discriminability mode is inconsistent with the finding that tasks with a primary dimension lowest in discriminability (CT in Set A and CC in Set B) were faster than the condensation tasks. When the dimensions that were lowest in discriminability were secondary and relevant, RTs were still faster than the condensation tasks. This result is not consistent with the sequential-by-primary mode.

If it is assumed that the component time distributions of the three dimensions were less than perfectly correlated and did not overlap, then the parallel contingent model would predict that there would be no difference in RTs in tasks in which the least discriminable dimension was relevant. Thus when the least discriminable dimension was secondary and relevant or primary in the contingent tasks, then RTs should have been equal to the RTs in the condensation tasks. In these cases, RTs would be determined by the dimension that was slowest to process. However, the condensation tasks were considerably slower than these contingent task RTs. In fact, even if there was some overlap in the component times, so that occasionally the condensation task would be limited by the more discriminable dimension, it would be difficult to produce a difference as large as 100 msec. between the condensation and these contingent tasks. It seems plausible, therefore, that the sequential contingent mode provides a more accurate description of *S*'s processing.

A mode that is consistent with all the above results is the sequential contingent mode. The contingent modes are also consistent with the finding that more errors were attributable to misidentification of the value of the relevant secondary dimension, rather than of the primary dimension, although this result is also consistent with the sequential-by-primary mode. Given that RTs were related to the discriminability of the primary dimension and that there were more primary errors than secondary errors, the failure to find any appreciable effect of primary dimension discriminability on secondary error rates is evidence against parallel contingent processing. The small difference that was found was in a direction opposite to that expected from the parallel model. The finding that repetition effects were related to the primary-secondary status of the dimensions is additional evidence for the functional significance of the contingent relations.

Some additional comment is required, however, concerning those between-task comparisons that were inconsistent with the discriminability orderings derived from the data of the two-choice and filtering tasks. In discussing differences in performance attributable to differences in the dimensions that were processed, the primary emphasis has been on the discriminability (physical similarity) of the dimensions as the factor responsible for these differences in RT. While the reversal in difficulty of the CC and the CT tasks between Sets A and B indicated that physical similarity was operative in establishing differences among the contingent tasks, it is likely that other factors were also operative in producing differences among the contingent tasks within a set.

Under the general rubric of "codability," two factors, presumably operating relatively independently of physical similarity, deserve discussion. The first is related to the role played (or reflected) by specific language habits on performance in these tasks. There was much consistency among *S*s in the order in which they identified the values of a dimensionalized pattern when they were asked to describe it. These orders typically corresponded to the most frequently occurring adjective orders in the language. Such natural language stereotypes are illustrated when people tend to say BIG RED HOUSE rather than RED BIG HOUSE. In describing the stimuli used in these experiments, *S*s would typically mention the value of the size dimension first, the value of the color dimension second, and the value of the tilt dimension last. Thus a stimulus was most often described as LARGE, GREEN, TWO O'CLOCK or SMALL, RED TEN O'CLOCK, and rarely as TWO O'CLOCK, LARGE, GREEN or RED, TEN O'CLOCK, SMALL. If these natural language orderings influenced the order in which dimensions were identified in these experiments, the sequence determined by the contingent mode of processing in the CS tasks, where size would be identified first, would be most consistent with the language sequence. The serial contingent mode of processing in the CT tasks would involve a sequence of identifications that would be most incompatible with the language sequence in that tilt was to be identified first. If the degree of correspondence between the adjectival sequence and the sequence determined by the contingent mode of processing determined the amount of interference (or facilitation) that would be produced from language habits (or some factor reflected by language habits), then the CT task would suffer maximal interference and the CS task the least interference. The CC task would be intermediate. This argument is supported by an examination of Fig. 5 and 6 which show that when size and tilt or color and tilt were the relevant dimensions in the CT tasks, RTs were longer than when these pairs of dimensions were relevant in the CS or CC tasks, respectively. When color and size were the relevant dimensions, the prediction of RTs on the basis of the adjectival ordering is that RTs would be longer in the CC than in the CS task. This is contrary to the results shown in Fig. 5 and 6, but a direct comparison in this case is difficult to make in that when size was the relevant secondary dimension in the CC task, the right hand was used, while in the CS task, when color was the relevant secondary dimension, the left hand was used.

The second factor arises from difficulty in utilizing different S–R assignments or differences in S–R compatibility (Fitts & Seeger, 1953). Specifically, in the high-compatibility CT tasks and the FC tasks, it was possible to "code" the tilt dimension as a pointer in that the lower part of the diameter was in the portion of the circle which corresponded to the hand that was to be used in responding. For example, in the ten o'clock position, the lower portion of the diameter was in the right part of the circle and in some tasks the diameter was in the 10 o'clock position when the right hand was to be used. Some *S*s who were assigned to these tasks reported that they employed such a coding scheme. It is possible that the extremely poor performance shown by *S*s in the FS and low-compatibility FT groups was a result of their inability to use such a pointing code in processing the information from the tilt dimension. This explanation can resolve two inconsistencies in the ordering of the tasks determined on the basis of the discriminability of the relevant dimensions and the proportion of the time that they were relevant: (*a*) In the low-compatibility contingent tasks of Set B, CT results in longer RTs than CC even though in the high-compatibility conditions of that set CT is faster than CC (Fig. 4); and (*b*) in Set A, the difference in discriminability between the size and color dimensions was slight, but a sizable difference occurred between the FC (in which size and tilt were relevant) and the FS (in which color and tilt were relevant) tasks (Fig. 4). In the former filtering task, the tilt dimension was assigned in the more compatible "pointing" fashion, while in the latter task no consistent pointing code could be used since each hand had both tilt values assigned to it. If only discriminability was operative, then the difference between the FC and FS tasks would have been less than the difference between the FC and FT tasks since there was a comparatively large difference between the discriminability of the color and tilt dimensions.

The evidence for serial contingent processing would appear to be at variance with the evidence for parallel processing of multidimensional stimuli reported in other studies (Biederman & Checkosky, 1970; Hawkins, 1969). One possible resolution of this discrepancy is that the serial versus parallel modes represent an option available to the processor. In the present experiment, the instructions and/or contingent correlations could have induced a serial mode. Another possibility, stressed by Garner and Felfoldy (1970), is that the manner of processing a dimensionalized set of stimuli may depend on the specific stimulus dimensions used to generate the set. Following Shepard (1964) and Lockhead (1966), they distinguish between integral (or unanalyzable) and nonintegral (or analyzable) stimulus dimensions, "the distinction phenomenologically being between dimensions which can be pulled apart, seen as unrelated, or analyzable, and those which cannot be analyzed but somehow are perceived as single dimensions. [Garner & Felfody, 1970, p. 225]." In a recent series of experiments, Garner and his associates (Felfoldy & Garner, 1971; Garner, 1970; Garner & Felfoldy, 1970) have argued that stimuli generated from integral dimensions typically yield results (e.g., redundancy gains) that have been interpreted as parallel processing, while nonintegral dimensions typically yield results that are consistent with serial processing. In the present experiment, the stimulus dimensions were most likely nonintegral. Clearly, the diameter tilt was analyzable from the circle size or the background color. Since the appearance of the color was that of the background square and not that of the circle, it too was analyzable from the circle. Given these stimulus dimensions, it is not unexpected, from Garner's (1970) position, that serial processing would occur. According to Garner's integrality hypothesis, however, contingent processing *should not* be evidenced when identifying stimuli generated from integral dimensions (e.g., sizes and colors of shapes). This prediction awaits future test.

The evidence supporting the serial contingent mode of processing bears some relevance to the taxonomic system proposed by Posner (1964, 1966). In Posner's (1966) system, an issue was whether transformations were performed on sometimes relevant information on trials when that information was not relevant. The shorter RTs in the contingent than in the condensation tasks may be regarded as evidence that such transformations were not (always) performed. However, the contingent tasks required more processing time than the filtering tasks. If transformations were not performed on the information from an irrelevant secondary dimension, then what is accounting for this additional time? One possibility is that time was required to "switch attention" from the primary to the secondary dimension. Such an explanation has been employed in the analysis of dichotic listening studies (Broadbent, 1958). Assume that this "switching time" is independent of the discriminability of the dimensions being processed. The difference in processing time between matched (for discriminability and responses) contingent and filtering tasks would be constant. It is, unfortunately, very difficult to employ such a "subtraction method" since, as discussed previously, discriminability might not have been the only factor determining differences in RTs. It is not surprising, therefore, that the relevant comparisons shown in Fig. 5 and 6 are not consistent on this point. Another possibility is that the sometimes relevant information, when it is not relevant, is "attenuated" but not filtered. Such an explanation has been advanced by Treisman (1964a, 1964b) to account for the effects of information from an irrelevant source during dichotic listening. Treisman's (1964a, 1964b) model would imply that stimulus identification in these tasks is dependent upon activation of one of six internal representations. These units correspond to the six stimulus values. The activation threshold of these units varies as a function of various task conditions and *S*'s experience. In the contingent tasks, the units corresponding to the values of the primary dimension would have the lowest threshold. When one of these primary units is activated, the thresholds for the units corresponding to the relevant secondary dimensions would be lowered. For example, the two primary units for the CC task in Set A corresponded to the red and green values. If the "red unit" were activated, the threshold for the two units corresponding to the tilt values would be lowered. If the "green unit" were activated, the thresholds for the units corresponding to the size values would be lowered. This notion of a sequential facilitation between units has been employed by Treisman in her analysis of dichotic listening to speech (1964a).

*Repetition effects.*—A change in the value of the primary dimension resulted in longer RTs than did a change in the value of the relevant secondary dimension. This confirms Shaffer's (1965) finding that repetition effects for rules were of greater magnitude than repetition effects for signals to which the rules were to be applied. The present data confirm Shaffer's interpretation, but without the uncertainty as to which dimension was "primary" and which "secondary."

The difference in the magnitudes of the repetition effects for changes in the values of the two types of dimensions cannot be attributed to differences in the kind of response changes that they required. It will be recalled that in the high-compatibility contingent tasks a primary value change required a hand change. However, since differences in the magnitude of repetition effects for primary value changes versus secondary value changes also occurred in the low-compatibility tasks, where primary value changes did not always entail hand changes, then differences in the response changes that were required cannot be responsible for the effect. Moreover, in the FT and FC filtering tasks, which were similar to the high-compatibility contingent tasks with respect to the confounding of primary and secondary value changes with hand and finger changes, repetition effects of different magnitudes for the two types of response changes did not occur. It would thus appear that this difference in the magnitude of the repetition effects for primary and secondary dimensions, or to use Shaffer's (1965) terminology, for rules and signals, is specific to these functional variables rather than to differences in response dimensions.

An additional finding was that a primary value change eliminated any concommitant repetition effects attributable to changes in the values of the secondary dimensions. In other words, when the primary dimension in a contingent task changed value, the increment in RT which occurred as a result of this change was unaffected by simultaneous changes in the values of the secondary dimensions.

The repetition effects were largely associated with stimulus variables. RTs were lengthened when an irrelevant dimension changed value in the absence of changes in the values of any of the other dimensions and, therefore, in the absence of response changes.[3] That these repetition effects occurred in the filtering tasks, where the irrelevant dimension was never relevant, and, moreover, the magnitudes of these repetition effects were associated with the discriminability of the irrelevant dimension that changed value, adds additional support to the interpretation that at least part of the repetition effect must be attributed to variables that have their effect prior to response selection.

The one task that showed hand-change repetition effects was the FS task. Unlike the FC and FT filtering tasks, changes in the value of the primary dimension in the FS task resulted in longer RTs than those obtained when the secondary dimension changed value. A possible explanation for this discrepancy is that the color dimension in the FS task was employed in a manner similiar to the way in which the primary dimensions were used in the contingent tasks. It was previously mentioned that in the FS task, *S* was unable to utilize a coding scheme for the tilt dimension such that the lower portion of the diameter would point to the appropriate hand. In the FS task, the coding was such that the value of the color dimension could be used to determine whether the upper or the lower portion of the diameter was pointing to the appropriate finger. Thus, when the color was green, the left hand was to be used and the lower portion of the diameter correctly pointed to the appropriate finger. When the color was red, the upper portion correctly pointed to the appropriate finger to be used on the right hand.

The finding that repetition effects can be attributed to stimulus and functional variations rather than response variations is not consistent with Williams' (1966) and Bertelson's (1963) response interpretation of the repetition effect. As such, the repetition effect holds promise of being a relatively sensitive dependent variable with which to study the effects of variations of task structure and input variables on information-processing behavior, as was implied by the recent work of Schvaneveldt and Chase (1969).

*Locus of effects in the information-processing sequence.*—The third general problem of this experiment was the specification of the locus of the use of contingencies in information handling. With the exception of the CT task in Set B, the low-compatibility conditions manifested similar orderings and repetition effects to the high-compatibility contingent tasks. If the utilization of the contingencies was dependent upon a direct selection of response dimensions, then their utility should have been reduced in the low-compatibility conditions and consequently the differences among the contingent tasks should have been decreased. That this did not obtain is evidence that these rules were being employed in stages prior to that involved in the selection of responses. The results of the compatibility variable were, therefore, consistent with the results of the repetition effect.

In ruling out response selection and execution effects as the locus of the contingent processing, caution must be exercised before effects are attributed to perceptual processes. Given the adjectival order and coding effects, and the independence in the error data of dimension discriminability and error type, it is quite likely that the contingencies were *not* affecting stimulus preprocessing or an initial read-in stage but, instead, were involved in a more central stage. While a number of paradigms are now available for the specification of effects at preprocessing or response selection stages (Sternberg, 1969), the middle ground is still relatively unexplored. It has been 14 yr. since Broadbent (1958) proposed his information-flow model and the difficulty still remains of conceptualizing the functions of the channel, although how information gets selected for it has come under intense investigation (Treisman, 1969). In fact, it is still not clear to what degree the channel should be conceptualized in terms of (*a*) a set of general stages (or transformations) whose mode of operation (but not duration) is relatively independent of the to-be-processed information and, (*b*) a set of events specific to the to-be-processed information.

## REFERENCES

BERTELSON, P. S–R relationships and reaction times to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, 1963, 65, 478–484.

---

[3] If the irrelevant dimensions are filtered out, how could their trial-to-trial variations affect RTs? One possibility is that the values of an irrelevant dimension are processed after a response is initiated. It is for this reason that the term "self-terminating" was not used in the introduction to refer to an alternative to exhaustive processing. Self-termination would imply that stimulus processing ceases once sufficient information is available for a correct response.

Biederman, I., & Checkosky, S. F. Processing redundant information. *Journal of Experimental Psychology*, 1970, **83**, 486–490.

Biederman, I., & Kaplan, R. Stimulus discriminability and S–R compatibility: Evidence for independent effects in choice reaction time. *Journal of Experimental Psychology*, 1970, **86**, 434–439.

Broadbent, D. E. *Perception and communication.* London: Pergamon Press, 1958.

Egeth, H. E. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1966, **1**, 245–252.

Egeth, H. E., & Smith, E. E. On the nature of errors in a choice reaction time task. *Psychonomic Science*, 1967, **8**, 345–346.

Felfoldy, G. L., & Garner, W. R. The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception & Psychophysics*, 1971, **9**, 289–292.

Fitts, P. M. Cognitive factors in information processing. *Journal of Experimental Psychology*, 1966, **71**, 849–857.

Fitts, P. M., & Biederman, I. S–R compatibility and information reduction. *Journal of Experimental Psychology*, 1965, **69**, 408–412.

Fitts, P. M., & Seeger, C. M. S–R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 1953, **46**, 199–210.

Garner, W. R. The stimulus in information processing. *American Psychologist*, 1970, **25**, 350–358.

Garner, W. R., & Felfoldy, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, **1**, 225–241.

Hawkins, H. L. Parallel processing in complex visual discrimination. *Perception & Psychophysics*, 1969, **5**, 56–64.

Hodge, M. H. The influence of irrelevant information upon complex visual perception. *Journal of Experimental Psychology*, 1959, **57**, 1–5.

Hunt, E. B. *Concept learning: An information processing problem.* New York: Wiley, 1962.

Lockhead, G. R. Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, 1966, **72**, 95–108.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

Montague, W. E. Effect of irrelevant information on a complex auditory-discrimination task. *Journal of Experimental Psychology*, 1965, **69**, 230–236.

Mylrea, K. C. A sixteen-unit projector with microsecond turn-on time. *American Journal of Psychology*, 1966, **79**, 314–317.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Nickerson, R. S. Binary classification reaction time: A review of some studies of human information-processing capabilities. (Tech. Rep. No. 2004) Cambridge, Mass.: Bolt Beranek & Newman, 1970.

Posner, M. I. Information reduction in the analyses of sequential tasks. *Psychological Review*, 1964, **71**, 491–504.

Posner, M. I. Components of skilled performance. *Science*, 1966, **152**, 1712–1718.

Reitman, W. R. *Cognition and thought: An information-processing approach.* New York: Wiley, 1965.

Schvaneveldt, R. W., & Chase, W. G. Sequential effects in choice reaction times. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

Shaffer, L. H. Choice reaction time with variable S–R mapping. *Journal of Experimental Psychology*, 1965, **70**, 284–288.

Shepard, R. N. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1964, **1**, 54–87.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica*, 1969, **30**, 276–315.

Trabasso, T., Rollins, H., & Shaughnessy, E. Storage and verification stages in processing concepts. *Cognitive Psychology*, 1971, **2**, 239–289.

Treisman, A. M. Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 449–459. (a)

Treisman, A. M. Selective attention in man. *British Medical Bulletin*, 1964, **20**, 12–16. (b)

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–299.

Williams, J. A. Sequential effects in disjunctive reaction time. *Journal of Experimental Psychology*, 1966, **71**, 665–672.

(Received October 23, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 239–244

# OBSERVING RESPONSES AND DECISION PROCESSES IN VIGILANCE[1]

MICHAEL J. GURALNICK[2]

*National Children's Center, Washington, D.C.*

Vigilance performance was investigated within the framework of a two-stage model. A direct measure of the first phase, the observing response (OR) stage, was obtained by requiring *S*s to key press to produce the stimulus display containing either a signal or noise. The second phase, the detection-decision stage, was analyzed using both conventional measures and those derived from the theory of signal detection (TSD). Specifically, the effects of time, signal difficulty, and payoffs were investigated. Results indicated that the frequency of ORs increased with time, while vigilance performance decreased. It was suggested that simple frequency of ORs is under the control of variables different from those controlling the decrement function. Reliable estimates of TSD parameters were obtained for the difficult signal condition and were discussed in relation to the OR stage. Payoffs had no effect.

In a recent review, Swets and Kristofferson (1970) have suggested that vigilance tasks be placed more within the context of psychophysical experiments, especially when indexes derived from the theory of signal detection (TSD) are of major interest. Psychophysical analysis within the TSD framework has revealed that one can obtain estimates of two independent parameters: *S*'s sensory acuity ($d'$) and *S*'s willingness to report a signal (his criterion, $\beta$) (Green & Swets, 1966). Psychophysical procedures are, of course, highly structured and explicitly designed to maximize *S*'s observation of the display on every trial. Vigilance tasks, on the other hand, are generally characterized, among other things, by the lack of an explicitly defined correspondence between *S*'s observing behavior and the occurrence of stimulus events. In fact, it is this latter relationship that is most often under investigation in vigilance tasks (Jerison, 1970).

A two-stage model of vigilance behavior proposed by Jerison and Pickett (1963), if directly applied, most closely approximates a psychophysical procedure. The first stage consists of *S*'s decision of whether or not to observe the display and is a function of the costs and values associated with the response outcomes, as well as the a priori probability of signal occurrence. If, as in a psychophysical task, an observing response (OR) occurs, then the second stage, the analysis of the sensory information or detection-decision phase, is presumed to follow. Although Jerison (1970) further indicates that a TSD analysis is appropriate for the second stage, he has continued to elaborate upon and emphasize the importance of the observing response stage (Jerison, 1970; Jerison, Pickett, & Stenson, 1965). Holland (1958) has also suggested that vigilance data can be most readily understood in terms of observing responses in that the probability of detecting a signal depends on the rate of emission of ORs. To study the frequency of observing responses, Holland devised a technique which requires *S*s to make an overt OR, generally a key press, in order to briefly illuminate a display, thereby permitting *S*s to determine whether a signal is present. Holland found that ORs, i.e., key presses or eye movements (Schroeder & Holland, 1968), are reinforced by the detection of suprathreshold signals, even in the absence of any other extrinsic reinforcement.

---

[1] This study was supported by Grant MH18873-01 from the National Institute of Mental Health, United States Public Health Service. The author wishes to thank David Weisman for assistance in data collection and J. Michael Walsh for help during various phases of this project. Thanks also to Jim Hays, who wrote the computer program for some of the data analysis. This research was conducted while the author was at American University.

[2] Requests for reprints should be sent to Michael J. Guralnick, National Children's Center, 6200 Second Street, N. W., Washington, D. C. 20011.

The present experiment was designed to study vigilance performance within the framework of the two-stage model. Specifically, by employing a variation of Holland's technique, OR rate and detection-decision responses (for both conventional measures such as hits and false alarms and the TSD indexes $d'$ and $\beta$) were investigated as a function of time, signal difficulty, and payoffs. Below-"threshold" and highly discriminable signals were presented to determine whether uncertainty as to the existence of a signal (given an OR) and the occurrence of false alarms (FAs) affect the distribution of ORs over time. To facilitate this evaluation, the occurrence of false alarms was made more likely by including a payoff matrix which increased the value of a hit and the cost of a miss. In addition, with respect to the payoff factor, the role of extrinsic reinforcement provided in the form of points associated with a financial reward for the possible response outcomes was investigated. Research involving this variable has produced ambiguous results for conventional measures (Weiner, 1969). With regard to TSD indexes, Davenport (1968, 1969) in both auditory and vibrotactile tasks, and Levine (1966) in an auditory task, found $d'$ to be invariant with time and costs for misses and false alarms. These experiments also indicated that $\beta$ was higher for high costs than for low costs and increased with time on task, whereas positive values associated with hits had no effect.

## METHOD

*Subjects.*—Twelve students, obtained from an introductory psychology *S* pool served as *Ss*. Each was paid $2.50/hr for participation. An additional incentive of $25.00 was paid to that *S* who achieved the highest point total.

*Apparatus.*—The stimuli consisted of 35-mm. slides of two vertical lines projected (Lafayette Model, KT-800) on a rear projection screen from an adjacent control room to the *S* room. Nonsignal or "noise" stimuli were defined as two vertical lines identical in height. A signal stimulus was one in which the upper segment of the right-hand member of the pair of lines was slightly longer than the left-hand member. For *S* to determine whether a trial contained a signal or noise, he needed only make a pair comparison on that trial.

The *S* was comfortably seated approximately 1.2 m. from the screen in a normally illuminated room. An easily grasped hand switch was placed in his left hand. Depression of this switch (OR key) briefly illuminated the display. Response buttons marked "yes" (indicating he believed a signal to be present) and "no" (indicating no signal present) were located on the extended right arm of the chair. White noise presented via a speaker (approximately 75 db.) from a Grason-Stadler (Model 901B) white-noise generator was used to mask extraneous sounds. All events were automatically controlled by the appropriate electromechanical programming and recording apparatus located in the adjacent control room.

*Procedure.*—Each *S* was scheduled for 10 sessions (only 1 session per day), with each session lasting 60 min. Care was taken to schedule each *S* at approximately the same time each day. Pilot investigations using the Yes–No (Y–N) psychophysical procedure (Green & Swets, 1966) provided the basis for the selection of two classes of stimuli: (*a*) An "easy" signal yielded 99.75% hits and no more than .2% false alarms. In practice, virtually all *Ss* achieved 100% hits and 0% false alarms under these alerted conditions for the easy signal. (*b*) A "difficult" signal was defined as any combination of hits and false alarms that generated a $d'$ value range of 1.5–2.5. In all sessions, the sequence of events required *Ss* to first depress the hand switch, which after a 1.5-sec. delay was followed by a 200-msec. exposure of the stimulus event. The *S* was then required to press either the Yes or No key. A Y–N response was required on every trial and the OR key remained inoperative until the response occurred.

The first two sessions were devoted to instructions, training, and practice within the Y–N psychophysical procedure. After *S* became familiar with the task requirements, the sequence of events, and the slide stimuli, he was informed of the possible outcomes of his behavior as related to the occurrence of a signal or noise event, including an explanation of hits, misses, false alarms, and correct rejections. This was accomplished by first presenting a card consisting of the four-celled, stimulus–response matrix and then relating the matrix to *Ss'* actual performance. Appropriately labeled feedback lights provided both information as to *Ss'* accuracy as well as demonstrating the response outcome possibilities. When this was clear, the payoff matrix was introduced. The *Ss* were told that we had found it worthwhile in the past to provide points for each correct response and to subtract points for each incorrect response and that a bonus of $25.00 would be given to the person with the highest point total obtained over the last eight sessions. Two payoff matrices were described. The first was balanced, providing one point for either a hit or correct rejection and subtracting one point for either a miss or a false alarm [1, 1, −1, −1]. The second payoff matrix provided 10 points for a hit and 1 point for a correct rejection and subtracted 10 points for a miss and 1 point for a false alarm [10, 1, −10, −1]. The *Ss* were given practice for all four combinations of payoff matrix and signal difficulty using the Y–N psychophysical procedure. Immediate feedback as to *Ss'* accuracy was provided and, although *Ss* could rest as often as necessary since they initiated each trial by depressing the OR key, frequent rest periods were scheduled by *E*. Signals were programmed to occur randomly on 50% of the trials. The *Ss* were given this information along with information regarding the payoff matrix currently in effect.

The remaining eight sessions were vigilance sessions consisting of two consecutive sessions each of the four combinations of payoff matrix and signal difficulty. The order of conditions was random with the restriction that each condition appeared an equal number of times at each position in the sequence. The *Ss* were permitted to press the OR key as often as they wanted and told to try to detect as many signals as possible, thereby earning as many points as possible. Of course, no feedback or scheduled rest periods were provided during these 60-min. vigilance sessions, although *Ss* were aware of the particular payoff matrix and signal level in effect for that session. Prior to each session, *Ss* were informed of total points earned in the previous session and given a brief practice period.

Each vigilance session was divided into three 20-min. segments for programming and analysis purposes. Ten signals occurred within each 20-min. time block. The signal was available on a "limited hold" for 5 sec. Any OR during that time resulted in the presentation of a signal as well as terminating the limited hold. If the signal was not observed during this 5-sec. period, it was replaced by a nonsignal slide. The signal schedule was obtained by selecting randomly from a rectangular distribution of intersignal intervals, with the restriction that no intersignal interval be more than 240 sec. (twice the average interval) or less than 20 sec. This schedule was repeated three times throughout the session. Only the data from the second session of each payoff-matrix signal level combination was used for analysis.

## RESULTS

Percent hit data were arc sine transformed and an analysis of variance carried out on the transformed scores. Hits, misses, false alarms, and correct rejections were calculated only for those occasions when an OR occurred. A "miss" could also occur by simply not making an OR within 5 sec.; however, this information was treated separately.

The percent hits decreased significantly over the three time blocks, $F\ (2,\ 22) = 7.26,\ p < .005$, as illustrated in Fig. 1. The effect of signal difficulty was also significant, $F\ (1,\ 11) = 81.71,\ p < .001$; however, neither the interactions nor the effects of the payoff variable reached significance.

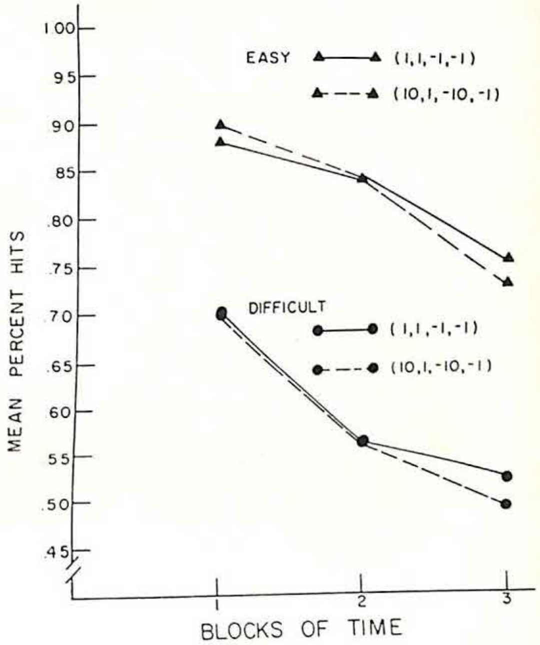


FIG. 1. Mean percentage of hits for each condition as a function of time.

For the Difficult Signal condition only, an analysis of variance was applied to the arc sine transformed percent false-alarm data. A significant decrease in the percent false alarms as a function of time was obtained, $F\ (2,\ 22) = 3.53,\ p < .05$, but no other effects were significant. For the Easy Signal condition, percent false-alarm data were analyzed using nonparametric tests (Siegel, 1956), since these data were markedly skewed. A Friedman two-way analysis of variance was carried out separately for each of the payoff conditions to assess the effects of time. Neither the [1, 1, −1, −1] nor the [10, 1, −10, −1] matrix produced a significant effect. A comparison between the two payoff conditions was also not significant (Wilcoxon test, $p > .05$). Finally, the percent false alarms for each *S* was much higher in the Difficult Signal condition than in the Easy Signal condition, with no overlap between the distributions.

The frequency of observing responses in-

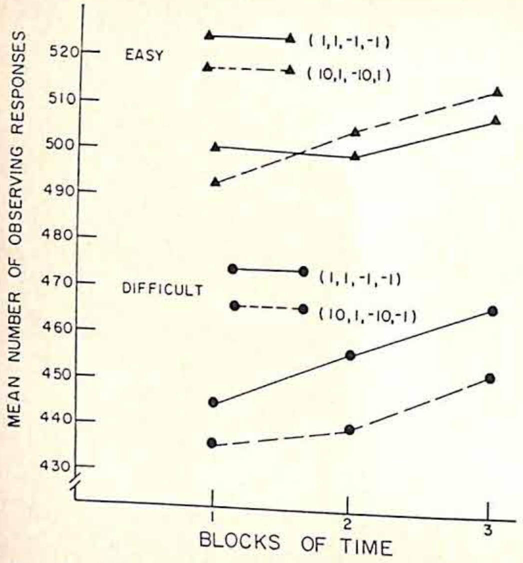


FIG. 2. Mean number of observing responses for each condition as a function of time.

creased significantly over time, $F$ (2, 22) = 7.32, $p < .005$ (see Fig. 2), and a substantially larger OR rate resulted during the Easy Signal conditions than during the Difficult Signal conditions, $F$ (1, 11) = 13.49, $p < .005$. The payoff variable again had no effect and none of the interactions approached significance.

The TSD measures, $d'$ and $\beta$ (see Table 1) were obtained from each condition from the hit and false-alarm data for each *S* for each of the three time blocks using Freeman's (1964) tables. However, TSD measures were derived for the Difficult Signal conditions only, for two reasons. First, the generally low rate of FAs that occurred in the Easy Signal condition does not permit reliable estimates of the detection theory parameters (Swets & Kristofferson, 1970). Second, a straightforward application of the TSD model to data including the frequent occurrence of either 100% hits and/or 0% FAs (as occurred in the Easy Signal conditions in this experiment) has been objected to on both theoretical and empirical grounds (Jerison, 1970; Jerison et al., 1965; Swets & Kristofferson, 1970). In the few instances in which this did occur in the Difficult Signal condition, an interpolation procedure following Loeb and Binford (1968) was employed.

Table 1

MEAN THEORY OF SIGNAL DETECTION (TSD) MEASURES AS A FUNCTION OF PAYOFF MATRIX AND TIME ON TASK

| Payoff matrix | Blocks of time | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| $(1,1,-1,-1)$: | | | |
| d′ | 2.11 | 1.77 | 1.75 |
| log $\beta$ | .3731 | .4521 | .5784 |
| $(10,1,-10,-1)$: | | | |
| d′ | 2.06 | 1.61 | 1.59 |
| log $\beta$ | .2548 | .3903 | .5819 |

An analysis of variance revealed that $d'$ decreased significantly over time, $F$ (2, 22) = 3.64, $p < .05$, with no other variable or interaction significant. A similar analysis carried out on the log $\beta$ values indicated a significant increase in those values over time, $F$ (2, 22) = 7.76, $p < .005$, also with no other variables or interactions approaching significance.

The OR rates, even under the Difficult Signal conditions, were sufficiently high that most signals were observed during the 5-sec. limited hold. Although more signals tended to be missed in the Difficult Signal condition, the difference was quite small.

## DISCUSSION

The frequency of observing responses increased with time, whereas vigilance performance decreased. For the Easy Signal condition, this latter fact is reflected by the progressive reduction in percent hits as a function of time (vigilance decrement), a result reported by a majority of such studies (see Jerison & Pickett, 1963; Mackworth, 1970). With respect to the increased rate of ORs, using a variation of Holland's technique in which ORs were defined as button presses to illuminate neon bulbs and the failure of a bulb to light constituting a signal, Broadbent (1963) found that the number of ORs increased with time, as did Holland (1958), at least for those *S*s in Holland's study who rarely missed a signal. In addition, most *S*s in the present study often observed the signal, as measured by key depressions, but failed to report its existence. Similar results have been reported by Broadbent and found in studies using eye fixations as the OR measure (Baker, 1960; Mackworth, Kaplan, & Metlay, 1964; Schroeder & Holland, 1968). Furthermore, as illustrated in Fig. 1, this tendency increased over time, an effect also noted by Broadbent and Schroeder and Holland.

All of the research discussed above employed signals readily detectable by an alert *O*. The present study has extended this work to include difficult signals, thereby permitting additional analyses using TSD indexes. For the Difficult Signal conditions, the percentage of hits decreased with time at a rate very similar to the Easy Signal conditions. However, this reduction in hits was not totally due to the fact that *S*s increased their criterion, $\beta$, since the sensitivity parameter, $d'$, also decreased as a function of time. The OR rate was substantially lower for the Difficult than for the Easy Signal condition, primarily due to the increased decision time required by *S*s during the Difficult Signal sessions.

The design of this study was formulated in terms of a two-stage model of vigilance behavior described earlier. Jerison has further suggested (Jerison, 1970; Jerison et al., 1965) that the observing stage can be classified into at least three types of observing behavior: (*a*) alert, (*b*) blurred, and (*c*) nonobservation, and that vigilance data, including TSD measures, mainly reflect different proportions of time *S* spends observing in each of these response modes. The technique employed in this study provided direct information with regard to the occurrence or nonoccurrence (nonobservation) of an OR. In addition, given that a key press produces some mode of observing, the decrease in $d'$ as the session progresses apparently reflects an increase in the proportion of blurred as opposed to alert observing (Jerison et al., 1965). In any event, at least for the Difficult Signal condition, it is assumed that the $d'$ parameter reliably reflects sensitivity changes, whether due to different observing modes or to other processes. Furthermore, criterion changes can be identified.

The use of an observable and independent measure of ORs here has provided data that suggest that, in this situation, the simple occurrence or nonoccurrence of an OR is under the control of variables different from those controlling the decrement function, since these two measures were unrelated. The similarity of the forms of the OR curves over time for all conditions suggests that the OR function is not closely related to the level of signal difficulty (with associated properties such as a feedback function for easy signals not totally shared by difficult signals), the percentage of false alarms, or other characteristics intrinsic to the signal detection situation. Rather, factors such as the overall payoff may determine the distribution of ORs over time. However, when signal detections are the only specified goal, as in tasks such as Holland employs (Holland, 1958; Schroeder & Holland, 1968), signal detections may indeed serve as reinforcers for ORs and control their frequency of occurrence.

The Easy Signal condition yielded an extremely low rate of false alarms with many *S*s not producing any at all. The closeness to the lower limit of 0% may have prevented the normally obtained result in vigilance studies, a decrease in false alarms as a function of time, from occurring. In the Difficult Signal condition, however, as usual, the percentage of FAs decreased and the criterion $\beta$ increased as a function of time. The general explanation for these changes is that *S*s have rather high expectations initially regarding signal probability and adjust their criterion appropriately as the session progresses (Colquhoun & Baddeley, 1964; Williges, 1969). The fact that the OR rate increased over time in this experiment suggests an additional possibility. As a consequence of the increased OR rate, *S*s in this experiment were experiencing increasing numbers of nonsignal trials, which effectively reduced the a priori probability of signal occurrence. This situation generally produces an increase in *S*'s response criterion (Baddeley & Colquhoun, 1969).

It was anticipated that by increasing the value of a hit and the cost of a miss, a lower $\beta$ value would result, if *S*s were attempting to maximize the expected value of their point total (Green, 1960). Although the trend was in the right direction, especially for the first two time periods, a statistically reliable effect was not obtained. In fact, the payoff variable did not have a detectable effect on any of the dependent variables. Perhaps applying costs directly to false alarms or increasing the amounts of differential costs and values will affect, not only TSD parameters, but the observing response stage as well.

## REFERENCES

BADDELEY, A. D., & COLQUHOUN, W. P. Signal probability and vigilance: A reappraisal of the "signal rate" effect. *British Journal of Psychology*, 1969, 60, 169–178.

BAKER, C. H. Observing behavior in a vigilance task. *Science*, 1960, **132**, 674–675.

BROADBENT, D. E. Some recent research from the Applied Psychology Research Unit, Cambridge. In D. N. Buckner & J. J. McGrath (Eds.), *Vigilance: A symposium.* New York: McGraw-Hill, 1963.

COLQUHOUN, W. P., & BADDELEY, A. D. Role of pretest expectancy in vigilance decrement. *Journal of Experimental Psychology*, 1964, **68**, 156–160.

DAVENPORT, W. G. Auditory vigilance: The effects of costs and values on signals. *Australian Journal of Psychology*, 1968, **20**, 213–218.

DAVENPORT, W. G. Vibrotactile vigilance: The effects of costs and values on signals. *Perception & Psychophysics*, 1969, **5**, 25–28.

FREEMAN, P. R. *Table of d' and beta.* (Applied Psychology Research Unit Rep. No. 529/64), Cambridge, England, 1964.

GREEN, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, 1960, **32**, 1189–1203.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

HOLLAND, J. G. Human vigilance. *Science*, 1958, **128**, 61–67.

JERISON, H. J. Vigilance: A paradigm and some physiological speculations. *Acta Psychologica*, 1970, **33**, 367–380.

JERISON, H. J., & PICKETT, R. M. Vigilance: A review and re-evaluation. *Human Factors*, 1963, **5**, 211–238.

JERISON, H. J., PICKETT, R. M., & STENSON, H. H. The elicited observing rate and decision processes in vigilance. *Human Factors*, 1965, **7**, 107–128.

LEVINE, J. M. The effects of values and costs on the detection and identification of signals in auditory vigilance. *Human Factors*, 1966, **8**, 525–537.

LOEB, M. & BINFORD, J. R. *Variation in performance on auditory and visual monitoring tasks as a function of signal and stimulus frequencies.* (Sensory Research Laboratory Tech. Rep.) Louisville: University of Louisville, 1968.

MACKWORTH, J. F. *Vigilance and attention.* Baltimore: Penguin, 1970.

MACKWORTH, N. H., KAPLAN, I. T., & METLAY, W. Eye movements during vigilance. *Perceptual and Motor Skills*, 1964, **18**, 397–402.

SCHROEDER, S. R., & HOLLAND, J. G. Operant control of eye-movements during human vigilance. *Science*, 1968, **161**, 292–293.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SWETS, J. A., & KRISTOFFERSON, A. B. Attention. *Annual Review of Psychology*, 1970, **21**, 339–366.

WEINER, E. Money and the monitor. *Perceptual and Motor Skills*, 1969, **29**, 627–634.

WILLIGES, R. C. Within-session criterion changes compared to an ideal observer criterion in a visual monitoring task. *Journal of Experimental Psychology.* 1969, **81**, 61–66.

(Received May 10, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 245–248

# ATTENTION DEMANDS OF MEMORY RETRIEVAL[1]

STEVEN W. KEELE[2]

*University of Oregon*

Typically many stimuli impinge on a person at one time. When the processing of one stimulus interferes with processing another, the processing is said to take attention. Such processing can be conceived as falling into two general subdivisions: (*a*) Sensory stimulation activates (retrieves) information stored in memory; (*b*) based upon the activated memory information, some operation, such as initiating a response, is performed. To determine whether the retrieval stage demands attention, two sources of information, form and color, were simultaneously presented. Only color was relevant to the key-pressing response. When the irrelevant form spelled a word, such as GLASS, no interference with processing the relevant color was found. Yet it was concluded that the irrelevant information had contacted memory because when the word meaning was changed, such as to GREEN, the time to respond to the color of ink increased. A discrimination of meaning depends on information stored in memory. The evidence supports the view that memory retrieval is a nonattentive process; more than one source of information can contact memory at a time.

People are limited in processing the many stimuli that they encounter. This limitation is particularly noticeable when two stimuli, each requiring a different response, occur at about the same time. One response is usually delayed, indicating that both signals cannot be handled at once (Smith, 1967). When performing one task interferes with performing another, the tasks are said to require attention. One aspect of understanding attention, therefore, is to identify the processes that interfere when performing two tasks (Posner & Keele, 1970).

Most tasks can be conceptualized as involving two subprocesses: (*a*) retrieval of information from memory and (*b*) operations performed upon the retrieved information. A signal may elicit the retrieval of an appropriate response, the name of the signal, or some other meaning. The retrieved information then may be operated upon by initiating a response, storing the information, combining it with other information (Keele, 1970), mental counting (Schvaneveldt & Staudenmayer, 1970), or some other operation. The limitation in processing information could occur at either or both of the two stages, memory retrieval or operations. The present study used an interference technique to determine the source of limitation.

Morgan and Alluisi (1967), Well (1971), and others found that at least under certain circumstances, irrelevant signals, which require no response, cause no increase in the time to respond to relevant signals requiring action. If it can be shown that the irrelevant information must have contacted memory even though it did not interfere with processing the relevant information, then it can be inferred that processing limitations are subsequent to memory retrieval.

One indication that irrelevant information contacts memory is the well-known Stroop effect (Jensen & Rohwer, 1966). The reaction time (RT) to name a color increases when the color prints out an irrelevant color word (e.g., the word GREEN printed in red ink; the correct response is red). One interpretation of the Stroop effect is that it is not during the memory retrieval stage that interference occurs. Rather, as a result of simultaneous memory retrievals, the two sources of in-

---

[1] The research was supported by the Advanced Research Projects Agency of the United States Department of Defense, monitored by the Air Force Office of Scientific Research under Contract F44620-67-C-0009 and National Science Foundation Grant GB21020.

[2] Requests for reprints should be sent to Steven W. Keele, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

formation, color and form, lead to conflicting responses. However, to be sure that it is not the retrieval stage per se that results in the Stroop effect, it needs to be demonstrated that irrelevant, *noncolor* words cause no interference. A discrimination of meaning cannot be made prior to memory contact.

An earlier study by Klein (1964) looked at verbal RTs to colors of ink with both irrelevant color words and noncolor words. A control condition used colored asterisks. When compared to the control, even the noncolor words lengthened RT to name the colors of ink. It can be argued, however, that the name of the noncolor word was retrieved at the same time as the name of the ink color, but the result led to conflicting verbal responses. The *S* was faced with selecting which of the two competing responses to elicit. This was essentially Klein's interpretation.

The current experiment attempted to eliminate response conflict caused by noncolor words by using key presses rather than verbal responses. Different colors of ink were assigned to different keys. It was expected that irrelevant color words would lead to conflicting key presses but other words would not, since other words would not tend to evoke any particular key press.

## METHOD

*Subjects.*—Twelve students from the University of Oregon employment service performed in two sessions. They were paid $1.50 per session.

*Materials and procedure.*—Each *S* was shown five different types of stimuli in four different colors: red, green, blue, and yellow-orange on a white background. When a stimulus was presented, *S* classified it on the basis of color by pressing one of four response keys operated by the middle and index fingers of the two hands. The five types of stimuli were: (*a*) color words—RED, GREEN, YELLOW, and BLUE; (b) noncolor words—RAT, GLASS, YANKEE, and BIRD; (*c*) scrambled letters from the color words—RDE, GNERE, YWOLEL, and BELU; (*d*) mixed letterlike forms (Gibson forms) matched in length with the color words (See Fig. 1); and (*e*) pure Gibson forms again matched for length with the four color words (See Fig. 1). This latter condition consisted of strings of forms that did not vary. The form of the stimuli was irrelevant with correct classification being determined solely by the color of ink. The fifth condition served as a control in which there were no irrelevant, varying forms. A constant letterlike form instead of color patches was used in the control so the salience of the colors would be the same as in the other conditions. The lengths of the strings were also matched with the color words to further control salience.

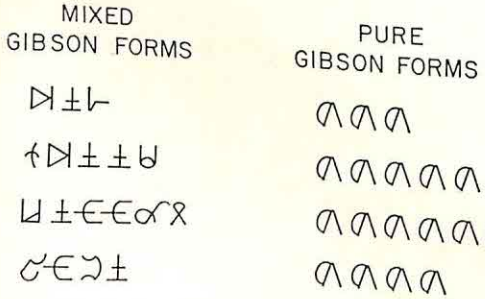


FIG. 1. Strings of forms used in the Mixed Gibson Form and Pure Gibson Form conditions. (The lengths of the strings match the length of the words red, green, yellow, and blue.)

TABLE 1

MEAN RT (IN MSEC.) AND PROPORTION ERRORS

| Measure | Pure Gibson Forms (control) | Mixed Gibson Forms | Scrambled letters | Noncolor words | Color words |
|---|---|---|---|---|---|
| RT | 554 | 553 | 560 | 559 | 604 |
| Errors | .08 | .09 | .10 | .07 | .09 |

The stimuli were tachistoscopically presented one at a time in a mixed order (determined by a random number table), and each of the five conditions were presented 36 times in each of the two sessions per *S*. Each stimulus presentation lasted for 1 sec. The latency of the key press classifying the color of each presentation was measured to the nearest millisecond and reported to *S* after each response. Approximately 8 sec. intervened between successive stimuli.

## RESULTS AND DISCUSSION

The mean correct RTs and proportion of errors for the five conditions of irrelevant information are presented in Table 1. The only condition showing a substantial RT difference from the control is the color-word condition. An analysis of variance performed on the RTs of all five conditions was significant at the .001 level of confidence, $F$ (4, 44) = 9.04. A subsequent analysis of RTs excluding the color-word condition was not significant, $F$ (3, 33) = .347.

The results suggest that even though

noncolor words did not interfere with responding to the color of ink, those irrelevant words must have contacted memory simultaneously with processing the color. The reason it is argued that the noncolor words contacted memory is that color words did interfere. Such a discrimination of word meanings could occur only at the level of memory. Thus, it is concluded that the limitation in processing multiple sources of information is not in the memory retrieval stage but subsequent to that stage.

A number of other studies support this conclusion that memory retrieval is nonattentive. Egeth, Blecker, & Kamlet (1969) required people to determine whether two colors were the same or different and found that the irrelevant words SAME or DIFF interfered with a same–different judgment of colors but other words did not. Their results are very similar to the present ones; the studies differ primarily in the implications that are drawn. Lewis (1970) has also shown that ignored messages in a shadowing task contact memory. Synonyms of words being shadowed interfere with the shadowing process when the synonym appears in the ignored message. The present study, along with the Egeth et al. and Lewis studies, indicates that *irrelevant* information contacts memory without delaying the processing of the relevant information. It is only when the irrelevant information leads to a conflicting response that interference arises.

Other studies involving more than one input with none of it irrelevant lead to a similar conclusion. Hawkins (1969) and Morton (1969) both found that combining two or more sources of information into a *single* response again produced no increase in processing time over a case with just one source of information. Hawkins had *S*s determine whether two stimuli varying on one, two, or three dimensions were the same or not. A single response, either same or different, was required regardless of the number of dimensions. The time to make the response when three dimensions were relevant was no longer than the slowest of the one-dimension conditions. In Morton's study, *S*s sorted cards into six piles. When each card contained two cues which were redundant and indicated the same sorting response, sorting time was faster than when only one cue was present. However, when two signals each lead to more than one response at a time (Karlin & Kestenbaum, 1968; Keele, 1970; and Schvaneveldt, 1969), interference does occur. The implication is that the requirement to make more than one response and not the requirement to process more than one signal is responsible for interference. Finally, when two successive letters are matched as being the same or different, no delay in RT occurs to a probe tone inserted shortly after the first of the letters (Posner & Boies, 1971). The lack of interference with the probe occurs even though the first letter is contacting memory at the time the probe occurs. Substantial interference with the probe RT occurs, however, once the second letter is presented and *S* initiates a response.

To summarize, information presented to the senses activates information stored in memory. Once memory information has been retrieved, various mental operations are possible, including initiation of a response. The present study is consistent with a number of others in suggesting that memory retrieval is not attention demanding; two or more sources of sensory information can activate information stored in memory simultaneously and without interference. Subsequent processes of selecting and operating on the activated memory information do appear to be attention demanding. If two operations are to be performed, such as two responses initiated, one or both responses are delayed. The only exceptions seem to be when there is no event and time uncertainty so that the responses can be emitted as a unit (Reynolds, 1966) and when there are extreme amounts of practice (Seibel, 1963).

## REFERENCES

EGETH, H. E., BLECKER, D. L., & KAMLET, A. S. Verbal interference in a perceptual comparison process. *Perception & Psychophysics*, 1969, 6, 355–356.

HAWKINS, H. L. Parallel processing in complex visual discrimination. *Perception & Psychophysics*, 1969, **5**, 56–64.

JENSEN, A. R., & ROHWER, W. D. The Stroop colour-word test: A review. *Acta Psychologica*, 1966, **25**, 36–93.

KARLIN, L., & KESTENBAUM, R. Effects of number of alternatives on the psychological refractory period. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 167–178.

KEELE, S. W. Effects of input and output modes on decision time. *Journal of Experimental Psychology*, 1970, **85**, 157–164.

KLEIN, G. S. Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 1964, **77**, 576–588.

LEWIS, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, **85**, 225–228.

MORGAN, B. B., JR., & ALLUISI, E. A. Effects of discriminability and irrelevant information on absolute judgments. *Perception & Psychophysics*, 1967, **2**, 54–58.

MORTON, J. The use of correlated stimulus information in card sorting. *Perception & Psychophysics*, 1969, **5**, 374–376.

POSNER, M. I., & BOIES, S. Components of attention. *Psychological Review*, 1971, **78**, 391–048.

POSNER, M. I., & KEELE, S. W. Time and space as measures of mental operations. Paper presented at the meeting of the American Psychological Association, Miami, September 1970.

REYNOLDS, D. Time and event uncertainity in unisensory reaction time. *Journal of Experimental Psychology*, 1966, **71**, 286–293.

SCHVANEVELDT, R. W. Effects of complexity in simultaneous reaction time tasks. *Journal of Experimental Psychology*, 1969, **81**, 289–296.

SCHVANEVELDT, R. W., & STAUDENMAYER, H. Mental arithmetic and the uncertainty effect in choice reaction time. *Journal of Experimental Psychology*, 1970, **85**, 111–117.

SEIBEL, R. Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology*, 1963, **66**, 215–226.

SMITH, M. C. Theories of the psychological refractory period. *Psychological Bulletin*, 1967, **67**, 202–213.

WELL, A. The influence of irrelevant information on speeded classification tasks. *Perception & Psychophysics*, 1971, **10**, 79–84.

(Received June 18, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 249-255

# ITEM ARRANGEMENT EFFECTS ON TRANSFER AND SERIAL POSITION ERRORS IN PART–WHOLE LEARNING OF DIFFERENT MATERIALS[1]

ALLAN L. FINGERET AND W. J. BROGDEN[2]

*University of Wisconsin, Madison*

Two lists (eight, two-digit numbers and eight CVC nouns) were learned in the part phase by alternate (AL) or successive (SUC) practice. Part-list items were interleaved for the double-alternation (DA) or the quadruple-alternation (QA) combination list practiced in the combination phase or throughout whole practice. Control *S*s for nonspecific transfer (NS) practiced DA and QA lists for which all items were different from those in the part phase. Positive transfer from part to combination phase was greater for specific than for NS transfer conditions. Form of serial position curve reflected the DA and QA patterns to a greater extent for specific transfer conditions than for NS or WH (whole-practice) conditions. Specific transfer part practice tended to be superior to WH practice. There were no significant differences between AL and SUC; but in part practice, the list of nouns was acquired more rapidly than the list of numbers.

Part–whole learning has usually been studied in terms of the efficiency of acquisition for a part-practice phase plus a combination-practice phase versus that for a whole-practice condition. Since the whole-practice condition and the combination phase of part learning must occur under identical conditions, a direct comparison of their efficiency of acquisition provides a measure of the transfer effect of prior part practice upon subsequent learning in the combination phase. In most studies of part–whole serial learning, the part lists were identified and acquired in a sequence during the part phase that was identical to the sequence by which the parts were combined to make the whole list. Although some systematic relationship between the whole list and its parts is necessary to make part–whole learning a meaningful process, relations other than ordinal sequence of fractions between part- and combination-practice phases may affect the transfer from part to combination learning and perhaps alter the relative efficiency of part versus whole acquisition.

In their study of part–whole learning, Fingeret and Brogden (1970) used two part lists, one of two-digit numbers and the other of CVC nouns. Use of two such lists provides the opportunity for construction of the combination or whole list by some systematic but unusual arrangements of the items of the part lists. If the two lists are interleaved to construct the combination list, that is, the first two items are the first two numbers (or nouns), the third and fourth items are the first two nouns (or numbers), the fifth and sixth items are the third and fourth numbers (or nouns), and so on, the arrangement of part-list items provides a pattern of double alternation (DA). If the two lists are interleaved by blocks of four items, that is, the first four items are the first four numbers (or nouns), the second four items are the first four nouns (or numbers), and so on, the combination list represents a quadruple alternation (QA) of the items of the two part lists. Of these two arrangements, transfer should be greater for QA than for DA because of the greater preservation of sequential item association in the QA combination list. The basic control group for measuring transfer from the part to the combination phases would be the whole learning of the DA and QA pattern. The arrangement of items of the two different

---

[1] This study was supported in part by a grant from the National Science Foundation.

[2] Requests for reprints should be sent to W. J. Brogden, Department of Psychology, University of Wisconsin, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706.

kinds of materials in DA and QA, however, may affect transfer from the part to the combination phase independently of the identity of the items. The possibility of nonspecific transfer requires control groups for which the noun and number items in the part phase are completely different from the items in the DA and QA combination lists.

Fingeret and Brogden (1970) found alternate trial practice of two lists (AL) to be superior to their successive practice (SUC) with mixed materials (one list of two-digit numbers and one of CVC nouns), but a reversed superiority when both lists were numbers or nouns. This result was interpreted to be due to the greater difficulty of learning two lists of numbers or nouns than of learning one list of numbers and one list of nouns. The arrangement of items from the part lists (one of numbers and one of nouns) in the DA and QA combination lists should increase the difficulty of acquisition in the combination phase over that for the combination lists used by Fingeret and Brogden for which the part lists of nouns and numbers were the first or second halves. Increasing the difficulty of the combination phase may alter the difficulty of AL and SUC part practice relative to each other and to whole practice.

In considering the relative efficiency of part versus whole practice for the DA and QA arrangements of items, it is dubious whether part learning will be superior to whole learning. Recent studies of part–whole serial learning by Postman and Goggin (1964), Goggin and Stokes (1969), Gladis and Abbey (1969), and Fingeret and Brogden (1970) have universally found part-practice procedures to be superior to whole-practice procedures. For all of these studies, however, the integrity of the part lists was maintained in making the combination or whole list. It may be hypothesized, therefore, that decreasing the identity of arrangement between items in the part-combination lists, would decrease the greater efficiency of part over whole practice.

The DA and QA arrangements of items from the two part lists may be hypothesized to affect the form of the serial position error curve for the combination list by an abrupt change in magnitude of error at serial positions involved in the change of kind of material. The change could be either an increase or decrease between the last item of a DA or QA series and the first item of the next DA or QA. Such an effect should be general and apply to all conditions of learning the combination or whole lists. With the specific transfer conditions, associative learning between items in the doublets and quadruplicates during the part phase may accentuate or alter the general affect hypothesized for pattern upon the serial position curve. The experiment to be reported was designed to test the above hypotheses.

## METHOD

*Subjects.*—The 160 *S*s were in classes of introductory psychology for which participation in experiments was a course requirement. They were assigned randomly in order of their appearance to the 10 treatments of the experiment ($n=16$ per group). All *S*s were tested by a single *E*.

*Design.*—There were three conditions of practice for each of two conditions of organization of material. The conditions of practice were two part methods, alternate (AL) and successive (SUC), and the whole (WH) method. Each part method involved learning two, eight-item lists, one of CVC nouns and one of two-digit numbers, each to a criterion of one errorless trial. With AL, the two lists were practiced on alternate trials until each was learned. With SUC, one list was practiced on successive trials to the criterial trial and then the other list was learned. The 16-item list of the combination phase was identical for both of the part methods and the WH method for each of the two organizations, double alternation (DA) and quadruple organization (QA). This list was practiced to a criterion of one errorless trial after completion of the part phase by *S*s in the AL and SUC conditions, and throughout practice by *S*s in the WH condition. The experimental groups were designated as follows: AL–DA, AL–QA, SUC–DA, SUC–QA, WH–DA, and WH–QA. There were four control groups for NS, one for each part method and each organization. The combination lists for these groups were identical to those for the experimental groups, but the items of the two part lists were different from the items used in the DA and QA list organization of the combination phase. The control groups were designated as follows: AL–DA–NS, AL–QA–NS, SUC–DA–NS, and SUC–QA–NS.

*Materials.*—The three-letter nouns (CVC trigrams) used in the construction of the eight-item lists were taken from Archer (1960) and had an association value in the 90%–100% range. Interitem similarity was minimized in list construction. The eight-item lists of two-digit numbers were constructed so that any given digit occurred no more than once in both the first and second ordinal positions. Neither zero nor double digits were used (e.g., 66). During the part phase, all *S*s in the AL and SUC practice conditions each learned a serial list of eight CVC nouns and a list of eight, two-digit numbers. There were two lists of CVC nouns and two lists of two-digit numbers. Half of the *S*s in each one of the AL and SUC conditions (DA, QA, DA-NS, and QA–NS) was given the sequence nouns–numbers in the part phase, and the other half was given the sequence numbers–nouns. Half of each of these two halves learned one pair of lists and the other half learned the other pair.

For the combination phase, the two, eight-item lists of the part phase were combined in either DA of QA orgainzation for all conditions except the NS conditions. In DA, the serial list of 16 items was composed of the first two nouns (or numbers) of Part List 1, then the first two numbers (or nouns) of Part List 2, then the third and fourth nouns (or numbers) of Part List 1, then the third and fourth numbers (or nouns) of Part List 2, and so on. In QA, the serial list was composed of the first four nouns (or numbers) of Part List 1, then the first four numbers (or nouns) of Part List 2, then the second four nouns (or numbers) or Part List 1, and finally, the second four numbers (or nouns) of Part List 2. For the NS conditions, the DA and QA arrangements were composed from the second pair of lists of nouns and numbers when the first pair was acquired in part practice and vice versa. The DA and QA organization of the lists used in the WH practice conditions were exact replicates of those used in the combination phase of the part-learning conditions. Counterbalancing of the two, eight-item lists of nouns and numbers and of noun–number sequencing was identical with that done for the part groups in the combination phase.

*Procedure.*—Slides of the items for the serial lists were projected black on a white background by a Kodak Carousel automatic projector on a "milk-plastic" screen in an aperture of the experimental chamber at a 2-sec. rate. The *S* was seated inside the experimental chamber at a table such that the projection screen was approximately 75 cm. from his face with vertical height of letters and digits at the screen of 2.5 cm. Slides of the letters A and B were used in the part phase to identify the beginning of the two lists on each trial for both the AL and SUC conditions. These letters also served as the anticipation cue for the first item. The letter A was used as the anticipation cue for the combination phase and for the WH condition. The interlist and intertrial interval was 10 sec. The interval between practice on List A and List B in SUC and the interval between the part and combination phases for AL and SUC was 30 sec. An intercom system between the experimental and control spaces permitted *E* to give instructions to *S* and to record the vocal anticipatory responses made by *S* in serial learning. Failures to respond and incorrect responses were recorded as errors.

For all treatment conditions, the instructions described the serial anticipation method and the performance required by the particular treatment condition. With the part methods, instructions were first given for the part phase, and no mention was made of the combination phase. When the criterial performance was attained, instructions were given for the combination phase. No specific information about DA or QA patterns was ever given. The *S*s in the AL conditions were given preliminary practice on two lists of two-letter words to ensure their understanding of the task. The *S*s in the SUC and WH conditions were given preliminary practice on a single list of two-letter words for the same purpose.

## RESULTS

The experimental design provides for statistical analyses of errors and time per item to the acquisition criteria. Analyses of variance were used in each case, and when appropriate Fisher's least significant difference test (LSD) of differences between means was made subsequent to analysis of variance. The 5% level was the criterion of significance for all statistical tests. Most analyses of the error data provided an evaluation of serial position and its interaction with other variables. Of the total errors, 75% were omissions (failures to respond) and 25% were intrusions. This proportional distribution of errors was relatively constant over experimental conditions, so no separate analyses were made.

*Combination phase analyses.*—Since the main hypotheses of the experiment were concerned with transfer effects of the part phase upon the combination phase, analyses were made of data from the WH groups and the combination phase of the part groups. In both the analysis of variance of errors and of time per item, groups was a significant source of variation, $F\,(9, 1{,}507) = 15.9$ and 13.0, respectively. Table 1 presents the mean errors and time per item for each group. With an LSD of 29.2 for errors and 6.9 sec. for time per item, each mean for specific transfer in the combination phase was significantly different and smaller than the corresponding mean for nonspecific transfer. All combination phase means were smaller and, except those for SUC–

TABLE 1

EFFECT OF PRACTICE CONDITION AND PATTERN OF MATERIAL UPON ERRORS AND TIME PER ITEM IN COMBINATION AND WHOLE ACQUISITION

| Practice cond. | Pattern of material | | | |
|---|---|---|---|---|
| | Double alternation | | Quadruple alternation | |
| | Specific | Non-specific | Specific | Non-specific |
| AL | 29.8(12.5) | 60.1(19.4) | 13.3(6.6) | 68.2(19.6) |
| SUC | 31.5(12.3) | 72.5(20.5) | 16.8(8.3) | 79.7(23.9) |
| WH | 139.8(34.6) | | 103.8(28.0) | |

Note.—Time per item is given in parentheses.

QA–NS, were significantly different from the corresponding means for the WH condition. There were no significant differences between AL and SUC practice for corresponding means. The only significant difference between corresponding means for the DA and QA conditions occurred with errors for WH practice.

*Serial position.*—The analysis of variance of the error data also showed the Groups × Serial Position interaction to be a significant source of variation, $F$ (135, 2,250) = 3.9. In order to evaluate curve form independently of differences in total error, the data for every $S$ at each serial position were converted to percentage of total error. Data for any $S$ who had five or fewer errors were omitted from the analysis of percentage of error because of the distortion they could produce. This procedure eliminated data for three *S*s from Group AL-DA, five *S*s from Group AL-QA, one $S$ from Group SUC-DA, and two *S*s from Group SUC-QA. The unequal $N$ analysis of variance of the percentage of error data showed the Groups × Serial Position interaction to be significant, $F$ (135, 2,085) = 3.67. Figure 1 presents serial position curves in the left section so that differences in form as a function of practice method and pattern of material may be examined. The LSD is 3.0% error. The effects of DA (upper two panels, left section) appeared by approximately the fifth serial position. At this serial position and thereafter percent error increased for the first item of the doublet and decreased for the second item. The amplitude of oscillation was greater for the two part methods with specific transfer than for WH. There were significant differences between points for either or both of the part curves from points for the WH curve at five serial positions (9, 11, 12, 13, and 14). The form of the curves for AL and SUC was very similar, and only at one serial position (12) did the difference between the two exceed the LSD. With nonspecific transfer (Panel 2, left section) there were no significant differences between AL, SUC, and WH at any serial position. The amplitude of oscillation was small and the differences between the peak and either the preceding or following points exceeded the LSD at four locations for AL-DA-NS, and at two locations each for SUC-DA-NS and WH-DA.

The effects of QA were much less clearcut than those for DA. They were greater for specific transfer with the two part methods (Panel 3, left section) than for all other QA conditions (Panel 4, left section). They involved progressive increase in percent error over the four items of a given class of material for the first two quadruplicates and progressive decrease in percent error over the items for the last two quadruplicates. The changes were not symmetrical within or between AL-QA or SUC-QA. Successive peak errors occurred at Serial Positions 5, 10, and 13 for AL-QA and at 4, 9, and 14 for SUC-QA. For both of these curves, there were a substantial number of points on the curve that were significantly different by the LSD test from the preceding and succeeding points. The two curves differed significantly from each other at 5 serial positions (5, 10, 11, 14, and 15). Either one of these two curves or both of them together differed significantly from the WH-QA curve at 10 serial positions (5, 6, 8, 9, 10, 11, 13, 14, 15, and 16). The curves for nonspecific transfer and WH practice (Panel 4, left section) show very little effect of the QA pattern. Only in the case of SUC-QA-NS was there a significant difference between adjacent serial positions (9 and 10). There were no significant

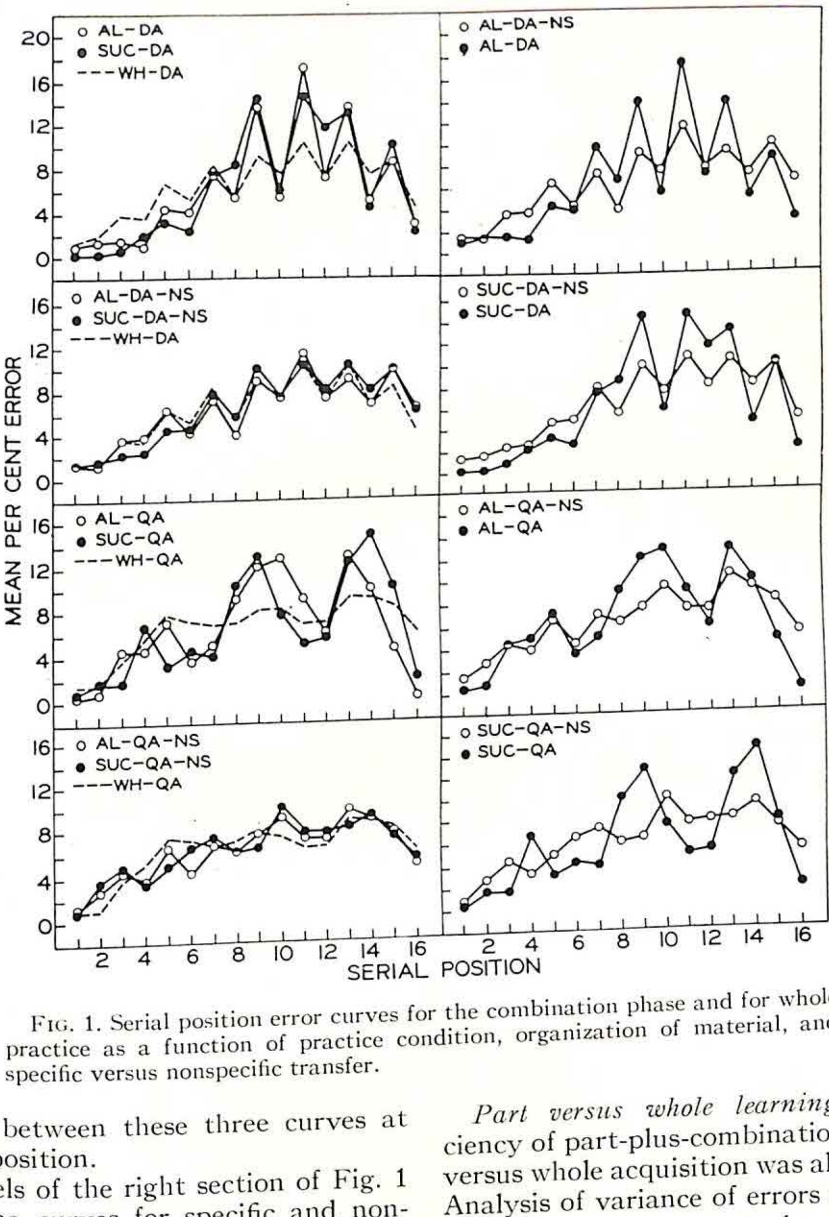


FIG. 1. Serial position error curves for the combination phase and for whole practice as a function of practice condition, organization of material, and specific versus nonspecific transfer.

differences between these three curves at any serial position.

The panels of the right section of Fig. 1 compare the curves for specific and nonspecific transfer for both AL and SUC practice and for both DA and QA organization. In each case, the effect of specific transfer and pattern of material on the form of the serial position curve was greater than the effect of pattern of material alone. Without detailing the serial positions, there were four significant differences between the curves in Panel 1, five in Panel 2, four in Panel 3, and seven in Panel 4. The contribution of specific items to the form of the serial position curve is substantial for both DA and QA organization.

*Part versus whole learning.*—The efficiency of part-plus-combination acquisition versus whole acquisition was also examined. Analysis of variance of errors and time per item showed groups to be a significant source of variation in both cases, $F$ (9, 150) = 3.7 and 2.6, respectively. Table 2 presents the mean errors and time per item for each group. All specific transfer part-practice conditions produced smaller means for both measures than the means for the respective whole-practice conditions, but with an LSD of 37.7 for errors and of 9.8 sec. for time per item, only the error means for AL–DA, SUC-DA, and SUC-QA differed significantly from the error means of the WH-DA and WH-QA groups.

TABLE 2

EFFECT OF PRACTICE CONDITION AND PATTERN OF MATERIAL UPON ERRORS AND TIME PER ITEM IN PART–PLUS–COMBINATION AND WHOLE ACQUISITION

| Practice cond. | Pattern of materials | | | |
|---|---|---|---|---|
| | Double alternation | | Quadruple alternation | |
| | Specific | Nonspecific | Specific | Nonspecific |
| AL | 87.1(28.0) | 103.6(32.4) | 75.7(22.3) | 123.1(32.4) |
| SUC | 86.1(27.1) | 115.8(32.4) | 61.2(20.0) | 130.6(37.9) |
| WH | 139.8(34.6) | | 103.8(28.0) | |

Note.—Time per item is given in parentheses. Abbreviations: AL = Alternate; SUC = successive; and WH = whole.

Comparisons of the NS part-practice conditions with WH practice showed a significant difference in mean time per item for SUC-QA-NS and WH-QA.

A further examination of data for the part and combination phases may be made by determination of the proportion of the total errors and total time per item contributed by each. Analyses of these relations were made by computing the percentage of the total errors and the total time per item for every *S* in the combination phase in each of the part groups and then conducting analyses of variance. Significant differences between groups were found for both measures, $F\ (7,\ 120) = 31.9$ and 16.2, respectively. Table 3 presents the group means for percent error and percent time per item. The LSD for percent error was 9.1 and for percent time per item was 9.0, and the results of the LSD tests were the same for both sets of data. The previous results reported for the combination phase were confirmed. There was significantly more transfer for the specific than for the nonspecific transfer conditions. That is, significantly greater percentages of errors and time were required in the combination phase for the nonspecific transfer conditions than for the specific conditions. The new result was a significant effect of pattern and practice method. The mean for AL-QA was smaller and significantly different from the means for AL-DA, SUC-DA, and SUC-QA.

*Part-phase analyses.*—The results of analyses of variance of the error data and the time per item data of the part phase were identical for the two measures. Neither practice method (AL vs. SUC) nor the dummy variables of pattern of organization (DA vs. QA) and specific versus nonspecific transfer were significant sources of variation. Ordinal position of list, $F\ (1,\ 124) = 19.9$ and 17.6, respectively; nouns versus numbers, $F\ (1,\ 124) = 23.8$ and 10.0, respectively; and the Practice Condition × Ordinal Position of List interaction, $F\ (1,\ 124) = 7.9$ and 7.8, respectively, were significant. Acquisition of the list of numbers produced more errors (29.8 vs. 21.6) and required more time (15.5 sec. vs. 12.6 sec.) than acquisition of the nouns. Application of the LSD test to the interaction means showed that only in the case of SUC practice was there a transfer from the first to the second list (30.5 vs. 17.8 errors and 8.1 vs. 5.1 sec.). The means for the first and second lists in AL practice (28.7 and 25.8 errors, and 7.7 and 7.1 sec.) and the mean for the first list in SUC practice (30.5 errors and 8.1 sec.) did not differ significantly.

## DISCUSSION

The most noteworthy result of the experiment was the effect of the DA and QA patterns of nouns and numbers upon the form of the serial position error curve. Each pattern produced a characteristic form that was present across all conditions of practice and transfer, with that for DA being somewhat more stable than that for QA. The magnitude of both pattern effects, however, was greater for the conditions of specific transfer than it

TABLE 3

EFFECT OF PRACTICE CONDITION AND PATTERN OF MATERIAL UPON PERCENTAGE OF ERRORS AND TIME PER ITEM IN THE COMBINATION PHASE

| Practice cond. | Pattern of material | | | |
|---|---|---|---|---|
| | Double alternation | | Quadruple alternation | |
| | Specific | Non-specific | Specific | Non-specific |
| AL | 30.0(42.0) | 59.3(57.5) | 15.4(26.1) | 55.6(54.1) |
| SUC | 32.8(42.1) | 59.2(60.4) | 25.2(40.0) | 62.3(63.4) |

Note.—Time per item is given in parentheses. Abbreviations: AL = alternate; SUC = successive.

was for the NS and WH conditions, which did not differ appreciably in magnitude. The occurrence of the same items in the two part lists and in the combination lists in the DA and QA patterns, therefore, increased the magnitude of the pattern effects. As hypothesized, this was probably due to the transfer of associative learning between items of the doublets and quadruplicates in the part phase to the combination phase.

The differential form of the serial position error curve between the DA and QA patterns must have been a function of pattern characteristics alone, since form for each pattern was relatively invariant across practice and transfer conditions. A basic characteristic of each pattern was the shift from one class of material to the other, and the points of shift were reflected in the form of the serial position curve. With the DA pattern, the increase in error for the first item of the doublet was undoubtedly produced by the shift in class of material and the decrease in error for the second item was probably produced by the repetition of an item of the same class. With the QA pattern, it is difficult to account for the progressive increases in error over the four items of the quadruplicate for the first two quadruplicates and the progressive decreases in error over the four items of the quadruplicate for the last two quadruplicates. The change in frequency of error between the last item of one quadruplicate and the first item of the next is undoubtedly due to the shift from one class of material to the other. The difference in direction of change in magnitude of error between the first two and last two quadruplicates may have been due to the QA pattern of the combination list functioning as four parts, or it may have been due to an interaction of ordinal position of the whole list with the QA pattern. Certainly, primacy effects appeared to override both DA and QA pattern effects, since the latter effects failed to occur or occurred to a lesser extent for the early ordinal positions than for the later ordinal positions. Since the standard serial position error curve shows an increase in error with ordinal position to just beyond the middle of the list and a decrease in error from this point to the end of the list, the difference in error progression over the quadruplicate between the first two and last two quadruplicates may be a simple reflection of this basic characteristic of all serial position error curves.

Although the expected positive transfer from part to combination learning was found for the specific transfer groups relative to both the NS and WH groups, the positive transfer for the NS groups relative to the WH groups was unexpected. Because there was no experimental plan to assess the sources of transfer for the NS groups, little can be said except to attribute it to warm-up, learning to learn, or some similar kind of variable. In addition it should be noted that there was little effect of type of part practice and type of pattern on transfer, except for the significantly smaller percentage of errors and time per item in the combination phase for AL-QA than for the other three specific transfer conditions. Thus, hypotheses about efficiency of transfer concerning either of these variables were not generally confirmed.

Since positive nonspecific transfer was found in comparing the combination phase to WH practice, it appears unlikely that systematic pattern arrangements other than those used in the present study would eliminate the superiority of part learning (with specific transfer) over whole learning. Goggin and Stokes (1969) have suggested that this superiority for serial learning is primarily due to advantages of the length-difficulty arrangement. In addition, variables of nonspecific transfer may also contribute to the superiority of part over whole learning. Furthermore, it appears that a wide range of variation in pattern relations between items in the part and combination lists may be added to the conditions under which part will be more efficient than whole serial learning.

## REFERENCES

Archer, E. J. Re-evaluation of meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

Fingeret, A. L., & Brogden, W. J. Part versus whole practice in the acquisition of serial lists as a function of class and organization of material. *Journal of Experimental Psychology*, 1970, **83**, 406–414.

Gladdis, M., & Abbey, O. Relationship between whole and part methods of learning and degree of meaningfulness of serial lists. *Journal of Experimental Psychology*, 1969, **81**, 194–196.

Goggin, J., & Stokes, C. Whole and part learning as a function of approximation to English. *Journal of Experimental Psychology*, 1969, **81**, 67–71.

Postman, L., & Goggin, J. Whole versus part learning of serial lists as a function of meaningfulness and intralist similarity. *Journal of Experimental Psychology*, 1964, **68**, 140–150.

(Received June 22, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 256–261

# VERBAL DISCRIMINATION LEARNING AS A FUNCTION OF PERCENTAGE OCCURRENCE OF REINFORCING INFORMATION (% ORI) AND VARYING PRESENTATION RATES

WILLIAM R. GAMBONI,[1] GREGORY R. GAUSTAD,[2]
AND BUFORD E. WILSON

*University of Nevada, Reno*

Two predictions derived from the frequency theory of verbal discrimination (VD) learning were that both reduced percentage occurrence of reinforcing information (% ORI) and an increase in presentation rate would retard VD learning. A third prediction was that % ORI would interact with presentation rate, resulting in a greater retardation of VD learning at the faster rate under reduced % ORI. Three levels of % ORI (100%, 50%, and 25% ORI) were combined factorially with two levels of presentation rate (2:1 and 2:4 sec.). Six independent groups of 15 college students each were used. All three predictions were confirmed.

According to the frequency theory of verbal discrimination (VD) learning (Ekstrand, Wallace, & Underwood, 1966), experimental frequency contributed by representational responses (RRs), pronunciation responses (PRs), and rehearsal-of-the-correct-alternative responses (RCRs) develops at a faster rate for the correct member of a VD pair compared to the incorrect member. Thus, a single VD list may be learned by an *S* consistently selecting the pair member of highest frequency (Rule 1). A minimal assumption of this theory is that an RCR occur at least once, and only when the incorrect member is selected.

Based on their theory, Ekstrand et al. (1966) were able to predict correctly an ease of learning order over several list conditions in which frequency was directly manipulated by using the same word twice within a single list and either maintaining or changing the initial designation (correct or incorrect) of the word.

Further support for frequency theory has come from studies in which transfer paradigms were used (Raskin, Boice, Rubel, & Clark, 1968; Smith & Jensen, 1971; Underwood, Jesse, & Ekstrand, 1964; Wallace & Nappe, 1970). Underwood et al. changed either the correct or incorrect words from VD List 1 to List 2, while Raskin et al. changed or maintained the correct and incorrect designation of words in two successive lists. Both Wallace and Nappe and Smith and Jensen gave familiarization training with correct and/or incorrect words later used in the construction of VD lists.

In addition to word repetition within lists and transfer paradigms, experimental frequency may be manipulated by the omission of information about the correctness of *S*'s choice for a number of VD pairs on a single trial (designated here as percentage occurrence of reinforcing information or % ORI).

Under conventional (100% ORI) VD learning, all selections are immediately followed by information indicating the correct member of the pair. Beginning with the initial guessing trial, chance would lead to the selection of the correct member in half of the pairs. Within these pairs, according to frequency theory, the correct member would gain two units of frequency (contributed by an RR and PR) compared to one unit of frequency (an RR) gained by the incorrect member. For pairs in which the incorrect member was selected, both correct and incorrect members would gain experimental frequency at the same rate (an RR and PR for the incorrect member and an RR and RCR for the correct member). The important point is that with VD learning under 100% ORI, learning theoretically begins on the second trial with the intrapair frequency relations either existing in favor of the correct member, or being equal. In no case would frequency relations be established in favor of the incorrect member of the pairs.

---

[1] Requests for reprints should be sent to William R. Gamboni, who is now at Chicago State University, Department of Psychology, Chicago, Illinois 60621.

[2] Now at Kansas State College of Pittsburg.

With VD learning under 50% ORI, frequency theory would lead to an expectation of a different pattern of intrapair frequency relations for certain pairs. As in the 100% ORI condition, chance expectancy would lead, more specifically, to four pairs correctly guessed on the first trial with an eight-pair list to be learned under 50% ORI. Of these four pairs, chance again would indicate that two would be followed by reinforcing information and two would not. This by itself does not constitute a critical difference in frequency outcome of members of correct pairs between 100% and 50% ORI since in both reinforced and unreinforced pairs the correct members would receive two units of frequency (an RR and PR) compared to one unit of frequency (an RR) for the incorrect members.

A critical difference between lists learned under 100% and 50% ORI would occur with respect to two of the four missed pairs in the 50% ORI list that would not by chance have been followed by reinforcing information. Within these two pairs, the incorrect member would receive two units of frequency (an RR and PR) and the correct member one unit of frequency (an RR), thus establishing frequency relations in favor of the incorrect member in these two pairs. The equalizing effect of an RCR would be precluded for these two pairs since no reinforcing information would have been available.

Based on these theoretical differences in frequency outcome between 100% and 50% ORI lists, it would be expected that VD learning under reduced % ORI would be retarded. Frequency theory (Ekstrand et al., 1966) also predicts that a reduction in study time after reinforcement would decrease the number of RCRs resulting in a retardation in VD learning. In the present study, it was also expected that a decrease in study time would retard VD learning. A third prediction was that presentation rate would interact with % ORI, resulting in a greater effect of rate on *S*s learning under reduced % ORI. This latter prediction stems from studies (Schulz & Runquist, 1960; Wilcoxon, Wilson, & Wise, 1961) which indicate that *S*s learning paired associates under reduced percentage of occurrence of response members (% ORM) tend to rehearse correct responses more than *S*s learning under 100% ORM. To the extent this same process is involved in VD learning, it would be expected that less rehearsal time would disproportionately affect learning under reduced % ORI compared to learning under 100% ORI. The purpose of the present experiment was to investigate VD learning under varying conditions of both % ORI and presentation rate.

## METHOD

*Design and subjects.*—The design of the present experiment was a 2 × 3 factorial, with 2 levels of presentation rate (2:1 and 2:4 sec.) and 3 levels of % ORI (100, 50, and 25% ORI). Six independent groups of 15 *S*s each were used. Groups 100–1, 50–1, and 25–1 were run under 100%, 50%, and 25% ORI, respectively, at the 2:1-sec. rate, and Groups 100–4, 50–4, and 25–4 were run under the same respective levels of % ORI at the 2:4-sec. rate.

The *S*s were 90 undergraduate summer session students with no previous experience in a verbal learning experiment. They were assigned randomly, within the constraint of equal *N*'s, to the six experimental conditions as each *S* appeared for testing.

*Lists and procedure.*—All *S*s were required to learn the same 24-pair VD list which varied among experimental conditions only in the number of pairs on a single trial that were followed by the correct member of the pair. The list was constructed with 48 words 5 and 6 letters in length randomly selected from a pool of 64 bisyllable nouns published by Spreen and Schulz (1966).

Under conventional (100% ORI) VD learning, all pairs were immediately followed by the presentation of the correct member on each trial. For the 50% ORI condition, only half of the pairs were accompanied by the correct member on any one trial. Pairs followed by the correct member were determined randomly within the limit that the correct member for all pairs occur twice in each block of four trials. Under 25% ORI, six pairs were followed by the correct member on a single trial. In this case, pairs followed by the correct member were randomly determined such that every pair was followed by the correct member once in each block of four trials. Under conditions of 50% and 25% ORI, no pair occurred on more than two or three consecutive trials, respectively, without the correct member being identified. When pairs were not followed by presentation of the correct member, an asterisk appeared in the window of the memory drum.

Four random orders of the list were constructed and presented serially on a Stowe memory drum. The members of a given pair were presented one above the other for 2 sec., followed by the presentation of the correct member or asterisk alone. For *S*s learning under the 2:1-sec. rate, the correct member or asterisk was presented alone for 1 sec. after the VD pair. For *S*s in the 2:4-sec. condition, the correct member or asterisk was presented alone for 4 sec. after each pair. In each of the four orders of the list, each member appeared equally often in the top and bottom positions and half of the top members in each list were correct. There was no intertrial interval.

All *S*s received conventional instructions about VD learning and, in addition, were told that the correct member may or may not be indicated for a given pair on any one trial. Learning was taken to the criterion of one perfect anticipation of the correct members in the 24 pairs on a single trial. If this criterion was not met within 23 trials, list presentation was discontinued.

## RESULTS AND DISCUSSION

The results were based on the mean number of trials to the learning criterion including the criterion trial. The *S*s who did not meet the criterion within 23 trials were credited with a criterion score of 23 trials. The number of *S*s who failed to reach criterion within 23 trials included four, two, and one *S*s in Groups 50-1, 25-1, and 25-4, respectively.

The mean number of trials to reach the learning criterion is presented in Fig. 1 for each of the six experimental groups. An analysis of variance of these data indicated the main effect of % ORI was significant, $F\ (2, 84) = 15.47, p < .001$. The overall means were 7.77, 13.04, and 12.77 trials to criterion for *S*s learning under 100%, 50%, and 25% ORI, respectively. This finding is consistent with the prediction derived from frequency theory that VD learning under reduced % ORI would be retarded. Figure 1 shows that for both the 2:1-sec. and 2:4-sec. presentation rates there was a general increase in the number of trials to criterion as % ORI decreased. The single exception to these trends was that Group 50-1, which included four *S*s who failed to meet criterion within 23 trials, took slightly more trials to criterion than Group 25-1.

The main effect of presentation rate was also significant, $F\ (1, 84) = 21.33, p < .001$. The mean trials to criterion was 9.18 and 13.20 for *S*s learning with the 2:4-sec. and 2:1-sec. rates, respectively. Thus, as predicted by frequency theory, a faster presentation rate, resulting in less study time, produced a retardation in VD learning. This finding is consistent with the results of studies by Ingison and Ekstrand (1970) and Wike and Wike (1970) in which the main effects of study time on VD learning were also significant.

Finally, the Rate × % ORI interaction was significant, $F\ (2, 84) = 4.93,\ p < .01$. As shown in Fig. 1, a reduction in presentation rate resulted in a greater retarding effect on VD learning under levels of reduced % ORI than under 100% ORI. Analysis of simple main effects of rate at each level of % ORI indicated that Group 50-1 took significantly more trials to criterion than Group 50-4, $F\ (1, 28) = 22.60$, $p < .001$, and Group 25-1 took significantly more trials to criterion than Group 25-4, $F\ (1, 28) = 4.86,\ p < .05$. Groups 100-1 and 100-4 did not differ significantly in the number of trials to criterion.

The significant Rate × % ORI interaction indicates that the trends of the curves presented in Fig. 1 are not the same. Since four *S*s in Group 50-1 failed to meet the learning criterion, the mean value for this group plotted in Fig. 1 was an underestimate and suggested the possibility of a quadratic trend component in the Rate × % ORI interaction.[3] Tests of both linear and quadratic components in the interaction term revealed that only the quadratic component was significant, $F(1, 84) = 6.51$, $p < .05$. Thus, the curves shown in Fig. 1 do not differ in slope; however, they do differ in curvature. Tests of linear and quadratic trends in the simple effects of % ORI at the 2:2-sec. rate of presentation showed that both the linear component (accounting for 69.76% of the % ORI effect) and the quadratic component (accounting for 30.24% of the % ORI effect) were significant, $F$ linear $(1, 40) = 20.88$, $p < .001$, and $F$ quadratic $(1, 40) = 9.05$, $p < .01$. Tests of trend at the 2:4-sec. rate showed that only the linear component (accounting for 97.84% of the % ORI effect) was significant, $F(1, 40) = 6.75$, $p < .05$. The present authors were unable to arrive at a tenable explanation of why the mean performance across levels of % ORI was linear at the 2:4-sec. rate and curvilinear at the 2:1-sec. rate. These findings would appear to be incompatible with the frequency theory of VD learning.

---

[3] The mean trials to criterion for Group 50–1 was 16.8. The median value for this group was 18.0, indicating the use of the mean resulted in a slight underestimate of the performance of Group 50–1.

The absence of an effect of presentation rate at the level of 100% ORI is not entirely inconsistent with the results of the studies by Ingison and Ekstrand (1970) and Wike and Wike (1970) mentioned above. In the study by Wike and Wike, individual comparisons revealed no significant differences in VD learning between groups of *S*s learning with 1-, 2-, and 4-sec. study intervals. A fourth group that received an 8-sec. study interval did differ significantly in the expected direction from the group that received the 1-sec. study interval. Similarly, in the study by Ingison and Ekstrand there was little difference in trials to criterion between *S*s learning with 2- and 3-sec. study intervals, 5.17 and 5.02 trials, respectively. However, *S*s that received a 6-sec. study interval took 3.88 trials to criterion. Although individual comparisons were not made by Ingison and Ekstrand, it appears that the main effect of study interval in their experiment was due primarily to the increased performance of *S*s who received the 6-sec. study interval. As in the present study, Skeen (1970) also found no effect of study interval on conventional (100% ORI) VD learning when study intervals of 2 and 4 sec. were used.

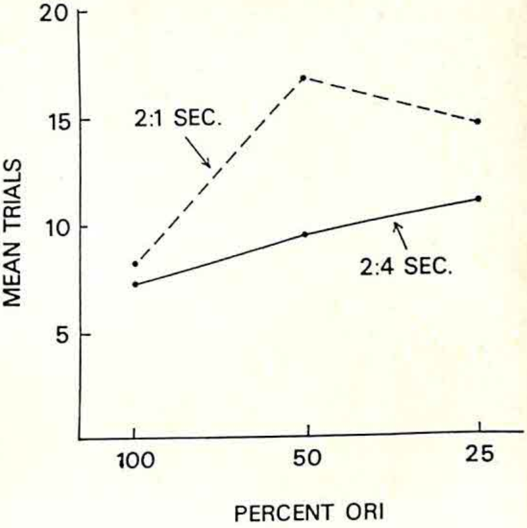


FIG. 1. Mean trials to criterion as a function of % ORI and presentation rate.

The finding that a reduction in presentation rate retarded learning under reduced % ORI but not under 100% ORI in the present experiment is compatible with observations by Schulz and Runquist (1960) and Wilcoxon et al. (1961), indicating that *S*s learning paired associates under reduced % ORM tend to rehearse more than *S*s receiving 100% ORM. In the present study, *S*s learning under conditions of 100%, 50%, and 25% ORI received an average of 7.77, 6.52, and 3.19 reinforcements, respectively, per VD pair during learning to criterion. These observations, similar to those of Wilcoxon et al., indicate that *S*s learning a VD task under reduced % ORI also receive greater benefit from each glimpse of the correct member of the VD pairs, perhaps mediated by greater rehearsal in these groups. According to frequency theory, the retarding effect of an increase in presentation rate on VD learning would occur as a result of fewer RCRs being made during the study interval. Thus, assuming *S*s in reduced % ORI conditions tend to rehearse more than *S*s in the 100% ORI condition, a reduction in presentation rate would have a greater effect of *S*s learning under conditions in which RCRs would be more important contributors to frequency differences between members of the VD pairs.

The implication of the argument presented above is that *S*s tend to compensate for reduction in % ORI with increases in rehearsal activity. Thus, it would be expected that a greater reduction in % ORI at a slower presentation rate would be necessary to produce a retarding effect on VD learning. In order to assess this possibility, specific between-group comparisons across % ORI conditions at each level of presentation rate were made using the Newman-Keuls test for multiple comparisons. These tests at the 2:1-sec. level of presentation rate indicated that both Groups 50-1 and 25-1 took significantly ($p$s $< .01$) more trials to criterion than Group 100-1, but that Groups 50-1 and 25-1 did not differ significantly on trials to criterion. Similar comparisons at the 2:4-sec. level of rate indicated the only significant ($p < .05$) difference was between Groups 100-4 and 25-4. These comparisons indicate that a greater reduction in % ORI was necessary at a slower presentation rate for % ORI to have a significant retarding effect on VD learning.

The finding that reduced % ORI retarded VD learning in the present experiment does not agree with the results of a study by Marston (1964) in which partial reinforcement as low as 50% during VD learning did not result in fewer trials to criterion compared to *S*s who received 100% reinforcement of correct responses. Although the reinforcement methodology used by Marston differed from the present study in that the onset of a green light served to indicate a correct response had been made, Marston did find that the number of reinforcements prior to criterion decreased significantly with decreases in percentage of reinforcement, a result similar to those obtained both by Wilcoxon et al. (1961) and the present study. Interpreted in the same manner that Wilcoxon et al. interpreted their results, this would mean that *S*s in Marston's reduced percentage reinforcement groups tended to rehearse more than *S*s receiving 100% reinforcement. Although Marston did not specify the presentation rate used in his study, if study time after reinforcement were relatively long, a greater reduction in percentage of reinforcement, as in the present study, would have been necessary to affect VD learning in Marston's study.

The equivalence of frequency outcome of members of correct pairs between reinforced and nonreinforced pairs discussed in the Introduction section was based on the accrual of experimental frequency according to the minimum requirements of frequency theory (Ekstrand et al., 1966) that an RCR be made at least once, and only when the incorrect member is selected. Nevertheless, it is reasonable to suggest that *S*s do rehearse when the correct member is selected and that rehearsal may involve more than a single RCR. With rehearsal after selection of the correct member, the frequency difference between correct and incorrect members would develop at a faster rate for reinforced pairs than nonreinforced pairs. In the case of reinforced pairs guessed correctly (CR pairs) on the initial guessing trial, the correct member would receive an RR, PR, and an indefinite number of RCRs compared to an RR for the incorrect member. The frequency ratio of members in CR pairs may then be represented as $2^+:1$ in favor of the correct member. The frequency ratio of members in correctly guessed but nonreinforced pairs (CN pairs) would remain 2:1 in favor of the correct member. Similarly, the frequency differences between members of incorrectly guessed pairs with reinforcement (IR pairs) would result in a frequency ratio of $2^+:2$ assuming indefinite RCRs, whereas the frequency ratio of members of incorrectly guessed nonreinforced (IN pairs) would remain 1:2 in favor of the incorrect member.

Based on these frequency ratios, it can be seen that CR pairs constitute a ratio most favorable to the correct member, followed by CN, IR, and IN pairs in decreasing order of favorableness. Of course, the intermediate positions in this ordering would depend on the amount of rehearsal of the correct member in IR pairs. Irrespective of the amount of rehearsal, providing its equal for both CR and IR pairs, it would be predicted that (*a*) a greater proportion of CR pairs would be correct on Trial 2 than either CN, IR, or IN pairs, (*b*) a greater proportion of CN pairs would be correct on Trial 2 than IN pairs, and (*c*) a greater proportion of IR pairs would be correct on Trial 2 than IN pairs.

In order to test these predictions, the proportions of CR, CN, IR, and IN pairs on Trial 1 that were correct on Trial 2 were determined for each *S* in the reduced % ORI groups. The *S*s in the two 100% ORI groups were excluded from this analysis since these *S*s did not experience CN and IN pairs. The mean proportions of CR, CN, IR, and IN pairs correct on Trial 2 were .65, .62, .61, and .44, respectively. The trend of these means is consistent with the order of favorableness of the frequency ratios for the correct member in each of the four types of pairs discussed above. A one-way analysis of variance for correlated measures of these proportion data indicated that the differences among the means were reliable, $F\ (3, 177) = 8.80$, $p < .001$. However, specific between-pair comparisons using the Newman-Keuls procedure revealed that the mean proportion of IN pairs correct on Trial 2 differed significantly ($p < .01$) from each of the remaining mean proportions, but that no other differences were significant.

These results lead to a number of suggestions. The finding that the mean proportion of CR pairs correct on Trial 2 did not differ significantly from the mean proportion of CN pairs suggests that *S*s do not rehearse appreciably after a correct member is selected. Second, the lack of difference between mean proportions of CN and IR pairs suggests that *S*s do rehearse when the incorrect member is selected, at least to the degree that the frequency ratio for IR pairs is commensurate to the frequency ratio for CN pairs. This would require a frequency ratio of 4:2 in favor of the correct member for IR pairs. Third, the fact that the mean proportion of IN pairs correct on Trial 2 was the only mean proportion less than .50 suggests *S*s were learning by selecting the pair member of highest frequency (i.e., Rule 1).

## REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Ingison, L. J., & Ekstrand, B. R. Effects of study time, method of presentation, word frequency, and word abstractness on verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **85**, 249–254.

Marston, A. R. Response strength and self-reinforcement. *Journal of Experimental Psychology*, 1964, **68**, 537–540.

Raskin, D. C., Boice, C., Rubel, E. W., & Clark, D. Transfer tests of the frequency theory of verbal discrimination learning. *Journal of Experimental Psychology*, 1968, **76**, 521–529.

Schulz, R. W., & Runquist, W. N. Learning and retention of paired adjectives as a function of percentage occurrence of response members. *Journal of Experimental Psychology*, 1960, **59**, 409–413.

Skeen, D. R. Study interval, manipulated frequency, and verbal discrimination learning. *Psychological Record*, 1970, **20**, 111–117.

Smith, S., & Jensen, L. Test of the frequency theory of verbal discrimination learning. *Journal of Experimental Psychology*, 1971, **87**, 46–51.

Spreen, O., & Schulz, R. W. Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 459–468.

Underwood, B. J., Jesse, F., & Ekstrand, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 183–186.

Wallace, W. P., & Nappe, G. W. Verbal-discrimination learning following a free-recall familiarization training procedure. *Canadian Journal of Psychology*, 1970, **24**, 27–33.

Wike, S. S., & Wike, E. L. Presentation rates and verbal discrimination learning. *Psychonomic Science*, 1970, **19**, 253–254.

Wilcoxon, H. C., Wilson, W. R., & Wise, D. A. Paired-associate learning as a function of percentage of occurrence of response members and other factors. *Journal of Experimental Psychology*, 1961, **61**, 283–289.

(Received June 29, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 262–267

# SEQUENTIAL DEPENDENCIES IN SINGLE-ITEM AND MULTIPLE-ITEM PROBABILITY LEARNING[1]

IRWIN P. LEVIN,[2] CORINNE S. DULBERG, J. FRANK DOOLEY,
AND JAMES V. HINRICHS

*University of Iowa*

A verbal discrimination task with 1, 2, or 4 letter pairs was presented where the letters within each pair were reinforced on a .75–.25 probabilistic reinforcement schedule and where reinforcement probability was conserved over 12 or 48 successive presentations of a given pair. Response tendencies dependent on the short-term sequencing of reinforcement events were shown to differ between groups. Analysis of variance tests sensitive to sequential dependencies yielded significance for 1-Pair and 2-Pair conditions but not for 4-Pair conditions. Results were discussed in terms of the reduced role of sequential information in multiple-item as compared to single-item probability learning.

The present study is a continuation of previous research in multiple-item probability learning (Erickson, 1966; Levin & Dooley, 1969, 1970; Voss, Thompson, & Keegan, 1959). Erickson has suggested that learning several simultaneous probability learning problems serves to interfere with *S*'s ability to use sequential information. The existence of sequential effects such as negative recency and response alternation led Anderson (1964) to question the validity of applying to the standard (single-item) probability learning task simple conditioning models (e.g., the stimulus sampling models described by Estes, 1964) which do not take into account response dependencies contingent on the sequencing of previous events. However, such conditioning models may be more applicable to multiple-item tasks with probabilistic reinforcement.

Levin and Dooley (1970), using a three-pair probabilistic verbal discrimination problem with reinforcement probability for a given pair conserved within each block of 10 successive presentations of that pair, found neither consistent positive recency, as would be predicted by the stimulus sampling models, nor negative recency; but the lack of systematic recency effects may indicate a reduction in the role of cues based on the sequencing of preceding reinforcement events. They did find evidence to suggest that some *S*s were slightly but reliably increasing their accuracy of prediction by using information about the frequency of event runs of different lengths, and thus that *S*s were being influenced to some extent by the sequencing of reinforcements. The present study used the Levin and Dooley probabilistic verbal discrimination paradigm and included manipulations that could influence the saliency of cues based on event sequences.

The major independent variable in the present study was the number of items (in this case, letter pairs) in the verbal discrimination list. The number of items included were one, two, and four for different groups of *S*s, with each item reinforced on a .75–.25 schedule. It was expected that as the number of items in the list was increased, the degree to which *S*s' responses depend on the exact sequencing of reinforcements for a given item would be decreased because of the increased difficulty in keeping track of event sequences for each individual item. The other independent variable manipulation was length of the reinforcement sequence. Jones and Myers (1966) showed that the length of the reinforcement block in probability learning—number of successive presentations of a given pair within which a given reinforcement ratio is conserved—is an important factor in the extent to which *S*s are influenced by event runs. In the present study, there were two different lengths of the reinforcement block for a given item: 12 and 48. Recency effects and other sequential dependencies were examined at each combination of the manipulated variables.

---

[1] The authors wish to express their thanks to David Reynolds, who developed the computer programs for data compilation and analysis. This project was supported in part by National Institute of Mental Health Grant MH-16362 awarded to James V. Hinrichs.

[2] Requests for reprints should be sent to Irwin P. Levin, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

## METHOD

*Design.*—A 3 × 2 between-*S*s design was used with 3 different list lengths and 2 different lengths of the reinforcement block. Lists of 1, 2, and 4 letter pairs were constructed. The 1-Pair lists were presented for 384 trials. The 2-Pair lists were presented for 288 trials per pair. The 2 pairs were interspersed randomly with the restriction that each pair appear equally often within each reinforcement block. The 4-Pair lists were presented for 192 trials per pair, with the 4 pairs interspersed randomly with the restriction that each pair appear equally often within each reinforcement block. A .75–.25 reinforcement schedule was maintained for each pair within each block of 12 or 48 presentations of that pair. For a reinforcement block of Length 12, one letter of a given pair was designated as correct on a randomly chosen 9 out of every 12 successive presentations of that pair and the other letter was designated as correct on the remaining 3 presentations. For a reinforcement block of Length 48, one letter of a given pair was designated as correct on a randomly chosen 36 out of every 48 successive presentations of that pair and the other letter was designated as correct on the remaining 12 presentations. Different reinforcement sequences were used for each different trial block within each half of the trials per session. The second-half sequences were repeats of the first-half sequences.

*Subjects.*—A total of 174 students from introductory psychology classes at the University of Iowa participated, with 29 *S*s assigned nonsystematically to each of the six experimental conditions. The *S*s were tested in groups ranging in size from 7 to 16 members.

*Materials.*—For the 1-Pair lists, the letter pair Q-V was used. For the 2-Pair lists the letter pairs K-W and B-X were used. For the 4-Pair lists the pairs Q-V, K-W, B-X, and J-Z were used. The eight different letters are of low frequency and have minimal formal similarity. Within each condition, subgroups were formed to counterbalance the assignment of a particular member of a letter pair to be the 75%-reinforced member of that pair. Across successive presentations of a given pair the right–left positions were randomized.

*Procedure.*—The *S*s were told that they would be presented a series of pairs of letters and that on each presentation they were to try to predict the correct letter of the pair. The correct letter was indicated by the presentation of a single letter following the presentation of the letter pair. Each *S* was seated in a booth before a response panel containing nine buttons. Prediction was made by pressing the button on either the far right or the far left of the panel according to the right–left position of the predicted letter within the letter pair. The sequence of letter pairs and single letters was presented on slides by a projector situated at the front of a room containing 20 individually boothed response panels. Presentation of the letter pair signaled the beginning of each trial. The presentation durations were 1.50 sec. for the letter pairs and 1.25 sec. for the single-letter feedback (reinforcement) presentations, with .75-sec. intervals between letter pair and single-letter presentations. Thus, *S*s had a 2.25-sec. interval between the onset of the letter-pair presentation and the onset of the single-letter presentation in which to make their response. Responses were automatically recorded on paper tape by a Microdyne Data Acquisition System.

The *S*s were given a 3-min. rest period midway through the experimental session, which lasted approximately 30, 45, or 60 min. depending on whether *S*s were in a 1-Pair, 2-Pair, or 4-Pair condition.

## RESULTS

*Mean learning curves and recency curves.*—Mean learning curves giving $P(A_1)$, the probability of responding with the 75%-reinforced member of a letter pair, as a function of trial blocks, are plotted in Fig. 1 for all conditions. It can be seen that an asymptotic level greater than the "probability-matching" level of .75 was reached in each condition. The asymptotic level tended to be slightly but insignificantly higher for 2-Pair and 4-Pair conditions than for 1-Pair conditions. For 1-Pair conditions, $P(A_1)$ tended to be higher with Reinforcement Block Length 12 than with Reinforcement Block Length 48, while the opposite was true for 2-Pair and 4-Pair conditions. However, none of the differences between reinforcement block lengths were statistically significant.

A recency curve—plot of $P(A_1)$ as a function of number of successive reinforcements of the most frequently reinforced response within a given pair ($E_1$ reinforcement events)—is plotted over the last half of the trials for each combination of the list length and reinforcement block

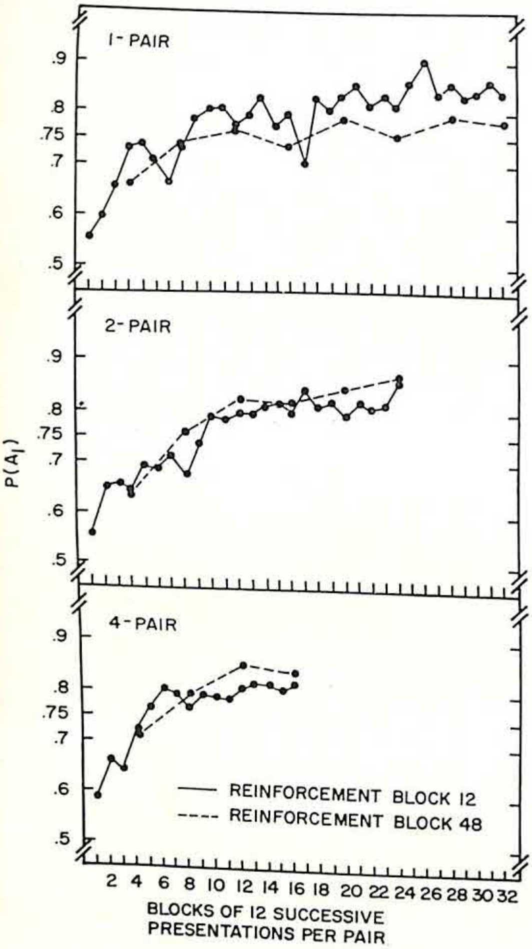


FIG. 1. Response probability as a function of trial blocks for each condition.

length in Fig. 2. For the 1-Pair condition with Reinforcement Block Length 12, $P(A_1)$ tended to increase slightly and then decrease as the number of successive $E_1$ events increased. The form of this curve corresponds to the negative recency often found in single-item probability learning. Some positive recency can be observed in the 2-Pair and 4-Pair conditions. However, none of the recency effects were large in magnitude or high in reliability.

*Sequential dependencies.*—Analyses of sequential dependencies were conducted within the framework of a stimulus averaging model applied to probability learning by Friedman, Carterette, and Anderson (1968). The linear and pattern models of probability learning (Estes, 1964) are special cases of this averaging model. This model, which considers the stimulus (reinforcement) events as pieces of information affecting choice behavior, has proven to be sensitive to response tendencies affected by short-term stimulus sequencing. The model may be written as

$$p_n = w_0 I_0 + \sum_{j=1}^{k} w_j s_{n-j},$$

where the $w_j$ are weights whose sum is unity, $p_n$ is the probability of an $A_1$ response on Trial $n$, $w_j$, and $s_{n-j}$ are the weight and value (taken as 0 or 1 for the two-choice task) of the event on Trial $n-j$, and $w_0$ and $I_0$ represent the weight and value of *S*'s expectation based on the sequence as a whole. It is assumed that $w_j$ is constant, independent of what events occurred on adjacent trials. This implies that in an analysis of variance of the $2^k$ possible reinforcement event subsequences for the $k$ previous trials, the interactions are zero. Response tendencies contingent on the exact sequencing of reinforcement events violate the independence-of-path assumption and lead to nonzero inter-

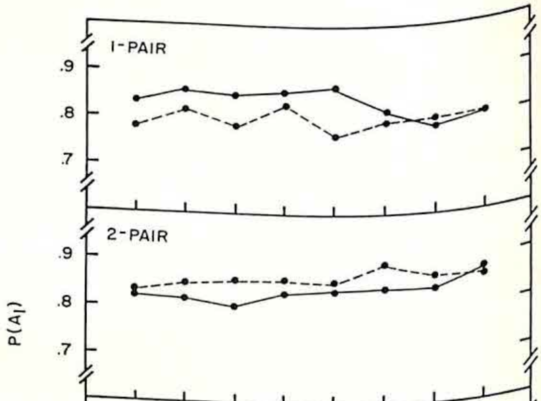


FIG. 2. Response probability as a function of number of successive $E_1$ events for each condition.

TABLE 1

Response Probability as a Function of the Reinforcement Event Sequence on the Preceding Three Trials

| Cond. | Event sequence ($E_{i,n-3}$, $E_{i,n-2}$, $E_{i,n-1}$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 111 | 121 | 211 | 221 | 112 | 122 | 212 | 222 |
| 1 Pair, Reinforcement Block Length 12 | .847 | .820 | .857 | .875 | .797 | .797 | .763 | .897 |
| 1 Pair, Reinforcement Block Length 48 | .789 | .749 | .823 | .833 | .692 | .655 | .791 | .448 |
| 2 Pairs | .849 | .815 | .834 | .837 | .778 | .746 | .786 | .664 |
| 4 Pairs | .842 | .814 | .830 | .807 | .784 | .742 | .765 | .752 |

Note.—Entries are probabilities of predicting the 75% -reinforced member of a given pair, conditional on occurrence of the specified event sequence for the previous three presentations of that pair.

actions. Thus the existence of significant interaction effects in any condition will show that the model is not applicable and that response choice in that condition depends on the sequencing of reinforcement events.

The model of Friedman et al. (1968) was applied to each of the list-length conditions by considering $P(A_1)$ for a given pair as a function of the eight possible event sequences of the previous three presentations of that pair and performing an analysis of variance over the last half of the trial blocks for each condition. The possible sources of variance include three main effects representing the degree of influence of each of the preceding three reinforcement events ($E_{i,n-3}$, $E_{i,n-2}$, $E_{i,n-1}$), three two-way interactions, and one three-way interaction. For the 1-Pair conditions, reinforcement block length interacted with other factors of the design, thus necessitating separate analyses at each reinforcement block length. For 2-Pair and 4-Pair conditions, no such interactions occurred and the two reinforcement block lengths were combined to increase power. The mean value of $P(A_1)$ for the eight event subsequences are given in Table 1 for each condition, and the analysis of variance results relevant for testing the model are presented in Table 2.

It can be seen that in each condition except 1 Pair with Reinforcement Block Length 12, the influence of events $E_{i,n-2}$ and $E_{i,n-1}$ were each statistically significant at or beyond the .01 level, with the value of $P(A_1)$ being greater following an $E_1$ on Trial $n - 2$ or $n - 1$ than following an $E_2$. For these conditions, the reinforce-

TABLE 2

Analysis of Variance Results for Event Dependencies of Table 1

| Source of variance | Cond. | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1-Pair Reinforcement Block Length 12 | | 1-Pair Reinforcement Block Length 48 | | 2 Pairs | | 4 Pairs | |
| | *MS* | $F(1,28)$ | *MS* | $F(1, 28)$ | *MS* | $F(1, 56)$ | *MS* | $F(1, 56)$ |
| $E_{i,n-3}$ | .062 | 2.58 | .000 | .00 | .033 | 1.35 | .005 | .48 |
| $E_{i,n-2}$ | .057 | 2.06 | .606 | 12.74** | .249 | 10.90** | .079 | 7.33** |
| $E_{i,n-1}$ | .076 | 2.07 | 1.337 | 13.49** | .941 | 27.44*** | .454 | 28.06*** |
| $E_{i,n-3} \times E_{i,n-2}$ | .116 | 5.29* | .238 | 6.76* | .021 | 1.01 | .009 | 1.19 |
| $E_{i,n-3} \times E_{i,n-1}$ | .000 | .00 | .185 | 4.08 | .047 | 2.20 | .001 | .06 |
| $E_{i,n-2} \times E_{i,n-1}$ | .075 | 3.81 | .443 | 6.33* | .111 | 4.79* | .000 | .02 |
| $E_{i,n-3} \times E_{i,n-2} \times E_{i,n-1}$ | .028 | 1.60 | .456 | 12.42** | .117 | 4.79* | .004 | .52 |

*$p < .05$.
**$p < .01$
***$p < .001$

ment event on the immediately preceding trial ($E_{i,n-1}$) was the largest contributor to the variance.

Inspection of Table 2 reveals that significant interactions were found in 1-Pair and 2-Pair conditions but not in the 4-Pair condition. Thus the stimulus averaging model is supported in the 4-Pair condition but not in the 1-Pair or 2-Pair conditions.

A close inspection of the means in Table 1 is useful in determining the nature of the interactions found in the 1-Pair and 2-Pair conditions. A disproportionately high value of $P(A_1)$ following Sequence 222 in the 1-Pair condition with Reinforcement Block Length 12 contributed to the significant interaction found in that condition. Since there were very few instances of runs of more than three consecutive $E_2$ events, this suggests that *S*s in this condition were effectively using information about the event run structure. The interactions found in the 1-Pair condition with Reinforcement Block Length 48 can be localized in two response tendencies. First, a response alternation tendency—the tendency for *S*s to predict the continuance of alternation following sequences such as 121 or 212—can be observed. $P(A_1)$ was higher following Sequence 212 than following any other sequence ending in an $E_2$ event, and $P(A_1)$ was lower following Sequence 121 than following any other sequence ending in an $E_1$ event. $P(A_1)$ was actually less following Sequence 121 than it was following Sequence 212. In addition, the value of $P(A_1)$ was disproportionately low following Sequence 222. Thus *S*s in this condition tended to predict a continuation in the run of $E_2$ events even though the actual probability of the run continuing was very small. Response tendencies similar to those observed in the 1-Pair condition with Reinforcement Block Length 48, but lesser in magnitude, can be observed in the 2-Pair condition. The observed interactions can be localized in response alternation tendencies following Sequences 121 and 212, and in a disproportionally low value of $P(A_1)$ following Sequence 222.

To reiterate the analysis of variance results for the 4-Pair condition, the influence of the reinforcement events on the previous two presentations of a given pair were each highly significant, but none of the interactions approached statistical significance. This indicates that while prior reinforcement events influence response choice on a given presentation, the exact sequencing of prior events does not affect response choice. These results support an averaging model which assumes that the weight or influence of a given event is independent of what events occurred on adjacent trials. This is in contrast to results in other conditions where rejection of the model was necessitated by the presence of sequential effects such as response alternation.

## DISCUSSION

Previous research has stressed the role of sequential dependencies in single-item probability learning. Thus, models based on the simple reinforcement principle that response probability should increase following every successive reinforcement of a given response tend to lose support in these studies where the effective stimulus on a given trial may be based on the sequencing of prior events. Several authors have suggested that choice tasks which separate the successive presentations of a given item serve to disrupt sequential information and the memory for preceding reinforcement events (Erickson, 1966; Kroll, 1970; Levin & Dooley, 1970; Millward, 1968). The present results for the 4-Pair condition tend to confirm this in the situation where successive presentations of a given probabilistically reinforced verbal discrimination pair are separated by the presentation of other pairs.

Assessed within the context of the Friedman et al. (1968) stimulus averaging model, response contingencies based on the short-term sequencing of successive reinforcement events for a given pair were found for 1-Pair and 2-Pair conditions but not for the 4-Pair condition. This represents the first direct statistical confirmation of the reduction of sequential effects in a multiple-item probability learning task. As the number of pairs in the list is increased, there appears to be a reduction in the saliency of cues based on the sequencing of reinforcement events.

Results in the 4-Pair condition were consistent with some of the predictions of a simple reinforcement theory. For example, the tendency to repeat the response designated as correct on the immediately preceding presentation of a given pair was the most important determinant of the present response. However, the prediction of probability matching generated by some models based on reinforcement principles (e.g., Estes, 1964) was not confirmed. Rather, the "overshooting" found in multiple-item probability learning problems by Erickson (1966) and Kroll (1970) was found here as well. While the sequencing of adjacent reinforcement events was not an important factor in the 4-Pair condition, response choice was affected by the type of event on each of the two previous presentations of a given pair. The results in this condition appear to illustrate the usefulness of the stimulus-integration approach to sequential processing represented by the averaging model.

## REFERENCES

ANDERSON, N. H. An evaluation of stimulus sampling theory: Comments on Professor Estes' paper. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

ERICKSON, J. R. On learning several simultaneous probability-learning problems. *Journal of Experimental Psychology*, 1966, **72**, 182–189.

ESTES, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

FRIEDMAN, M. P., CARTERETTE, E. C., & ANDERSON, N. H. Long-term probability learning with a random schedule of reinforcement. *Journal of Experimental Psychology*, 1968, **78**, 442–455.

JONES, M. R., & MAYERS, J. L. A comparison of two methods of event randomization in probability learning. *Journal of Experimental Psychology*, 1966, **72**, 909–911.

KROLL, N. E. A. Learning of several simultaneous probability learning problems as a function of overall probability and prior knowledge. *Journal of Experimental Psychology*, 1970, **83**, 209–215.

LEVIN, I. P., & DOOLEY, J. F. Response probabilities in a verbal discrimination task with probabilistic reinforcement. *Psychonomic Science*, 1969, **14**, 167–168.

LEVIN, I. P., & DOOLEY, J. F. Performance in a verbal discrimination task with items differing in reinforcement probability. *Journal of Experimental Psychology*, 1970, **84**, 508–513.

MILLWARD, R. B. Probabilistic reinforcement of reversal and dimensional shifts. *Journal of Mathematical Psychology*, 1968, **5**, 196–223.

VOSS, J. F., THOMPSON, C. P., & KEEGAN, J. H. Acquisition of probabilistic paired associates as a function of S-$R_1$, S-$R_2$ probability. *Journal of Experimental Psychology*, 1959, **58**, 390–399.

(Received July 3, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 268–278

# EFFECTS OF VERBAL CODING ON RECOGNITION MEMORY[1]

JOHN L. SANTA[2] AND HOWARD B. RANKEN

*Purdue University*

Four experiments investigated the role of verbal coding on retention of nonsense shapes as a function of memory load and retention interval. In training, *S*s in the Named and Unnamed conditions had equal practice discriminating shapes in a delayed matching-to-sample task, while *S*s in the Named condition also learned names for the shapes. A recognition test of the shapes followed pretraining. The Named condition, in general, was superior in terms of recognition performance. The advantage increased as the number of shapes to be remembered increased, but was constant across retention intervals from .2 to 50 sec. The superiority of the Named condition can be reduced or eliminated by interfering with rehearsal (counting backwards). Furthermore, estimates of shape recency in a continuous recognition task were superior for the Named condition only on items within the range of short-term memory. These data suggest that verbal coding of visual stimuli has its primary effect on recognition memory via increased rehearsal efficiency in short-term memory.

To what extent is memory of a picture or visual stimulus dependent upon a verbal process? This basic question has generated a long history of research interest resulting in some understanding of the role of verbal coding in memory. However, efforts to fully understand memory of visual materials have been somewhat retarded by the difficulty in generalizing across different task paradigms.

A variety of data suggest that prior label training facilitates performance on associative tasks where *S* must make new responses to visual stimuli or learn to pair nonsense shapes with each other (Ellis, 1965; Hagen & Kingsley, 1968; Jeffrey & Bogartz, 1961; Ranken, 1963; Santa & Ranken, 1968; Spiker, 1956). The effect of verbal coding is not as clear in other paradigms such as recognition memory of visual stimuli. Several investigators have found little or no effect of label pretraining in recognition tasks (Ellis, Bessemer, Devine, & Trafton, 1962; Ellis & Muller (6-point shapes), 1964; Hake & Erikson, 1956; Robinson, 1955). Other experiments have shown that prior label training facilitates recognition performance (Arnoult, 1956; Ellis & Muller (24-point shapes), 1964).

Efforts to resolve such conflicting data indicate that the affect of label training in a recognition task interacts with the complexity of the visual stimulus such that the labels are more important with complex stimuli (Ellis & Muller, 1964).

Studies of color codability suggest that the effect of verbal coding might also interact with other factors such as memory load and retention interval. In general, the codability literature demonstrates that the ease of naming visual stimuli is positively related to recognition performance (Brown & Lenneberg, 1954; Glanzer & Clark, 1962, 1963; Lantz & Stefflre, 1964). Stimuli that are readily coded with verbal labels are better remembered than those which are difficult to code verbally. The relationship is qualified such that verbal coding seems to be more important as either retention interval or memory load is increased (Brown & Lenneberg, 1954). However, the two variables of memory load and retention interval were partially confounded in the Brown & Lenneberg experiment, and the correlational data are subject to a noncausal interpretation. Namely, it is possible that stimuli which are easier to code are also more easily remembered, independently of the verbal code.

The first experiment in the present series was designed to examine experimentally the suggested interaction of verbal coding with memory load and retention interval.

---

[1] This research was based upon a doctoral dissertation submitted to Purdue University by the first author. The research was supported by a National Institute of Mental Health predoctoral fellowship.

[2] Requests for reprints should be sent to John L. Santa, Department of Psychology, Rutgers University, Douglass College, New Brunswick, New Jersey 08901.

## EXPERIMENT I

### *Method*

*Stimuli.*—The stimulus shapes (Fig. 1) were 10-sided nonsense shapes constructed with straight sides and irregular top and bottom contours. Thus, each shape could be divided by a horizontal line into a top and bottom half, each half containing one of the two contours which defined the shape. The shapes were projected on a rear-projection screen to appear 1 in. in width at a viewing distance of about 20 in.

*Apparatus.*—Presentation and retention intervals for the first experiment were controlled by an eight-bank timer (Lafayette Instrument Co.). Responses were recorded on paper tape with an Invac paper tape punch. A solid-state programming unit (Lafayette Instrument Co.) was used to program cue sequences, time intervals, and record responses in all succeeding experiments.

*Design*—Two training conditions, Named and Unnamed, 8 retention intervals (.2 to 50 sec.), and 3 memory loads (one, two, or four shapes) were used in a 2 × 8 × 3 design with repeated measures on the retention interval factor. A total of 60 *Ss* (introductory psychology students) participated in the experiment. The *Ss* were trained and tested individually. During training, half of the *Ss* were assigned to the Named condition, where they learned names for the shapes, while *Ss* in the Unnamed condition had equal practice in discriminating both top and bottom contours of the shapes in a delayed matching-to-sample task, but were not given names. Each of these training conditions was subdivided into three memory load groups for the recognition test phase. Thus, after training, *Ss* were assigned a fixed memory load for the test stage of one, two, or four shapes.

*Procedure.*—The shapes were first presented singly for 5 sec. each. The *Ss* in both Named and Unnamed conditions were informed that in a subsequent task they would be required to pay close attention to the whole shape, both top and bottom contours. In the Named condition, *Ss* were also given a name for each shape. The names were arbitrarily chosen, but were somewhat suggested by the outline of the shape (see Fig. 1).

Training trials were then given, in which a complete test shape was shown for 2 sec., followed, after a 4-sec. delay, by a set of four top or four bottom contours (half shapes), presented side by side for 3 sec. One of the four alternatives always matched the corresponding part of the test shape, and *S* was instructed to designate that alternative by pressing one of four response keys located on a table directly below the screen and in front of *S*. The *S* was to press the key corresponding to the position of the half shape which he chose. The *Ss* in the Named condition gave the name of the test shape when it was shown, with correction by *E* after 3 sec. The intertrial interval was 4 sec. The choice response, designating the matching contour, could be made at any time prior to the presentation of the next test stimulus, but was not corrected in either condition. A warning light mounted in the response key panel appeared 1 sec. before the presentation of the next test stimulus. Training continued for eight randomly ordered runs through the eight stimulus shapes. The sequence of half-shape alternatives was constructed such that within 32 recognition trials each stimulus was tested four times against four different sets of half shapes. The correct alternative for a given test stimulus appeared once in each of the four positions. The three distractor contours were randomly selected from the seven remaining stimuli. Each stimulus was tested twice against a set of top contours and twice against a set of bottom contours within a block of 32 recognition trials. Half shapes were used as alternatives in training to force *Ss* in Named and Unnamed conditions to attend equally to both of the relevant aspects of the stimulus shapes.

In the test task, *S* was presented with a memory set of one, two, or four shapes. The *S* was to remember the entire memory set over the retention interval which varied from .2 to 50 sec. Presentation times for memory sets of one, two, and four were, respectively, .5, 1, and 2 sec. However, in all cases, *S* was tested on only one member of the memory set. At the end of the retention interval, one of the four cue lights, located below the screen, came on for .5 sec. to indicate the position in the memory set which would be tested. The cue essentially told *S* that he would be tested on the shape which had been presented in Position $n$ ($n = 1, 2, 3$, or 4) of the memory set. As soon as the cue lamp went off, four whole-shape alternatives, selected from the eight stimulus shapes, appeared side by side on the screen for 4 sec., and *S* was to select the alternative that was the same as the item cued in the memory set. The three distractors were chosen from stimulus shapes not in the current memory set. The interval between the test alternatives and the next memory set was 4 sec., including a warning light after 3 sec. Memory load could be varied with this technique, and yet in all cases *Ss* were required to recognize only one shape from four alternatives. The cue was irrelevant for *Ss* in the condition with a memory load of one shape, but in this case the middle two lamps were lighted immediately preceding the test alternatives. These lamps were included to equate the total retention intervals and to provide a distraction similar to that experienced by *Ss* in the other conditions. When the memory load was two shapes, one of the two middle lights was presented on each trial to indicate the test position within the memory set and, when the memory set was four shapes, one of the four lamps indicated the test position.

The test sequence for a memory set of one was constructed so that there were eight randomly ordered runs through the set of eight stimuli. Within each block of eight recognition trials, the eight retention intervals (.2, 1, 2, 4, 8, 15, 30, and 50 sec.)

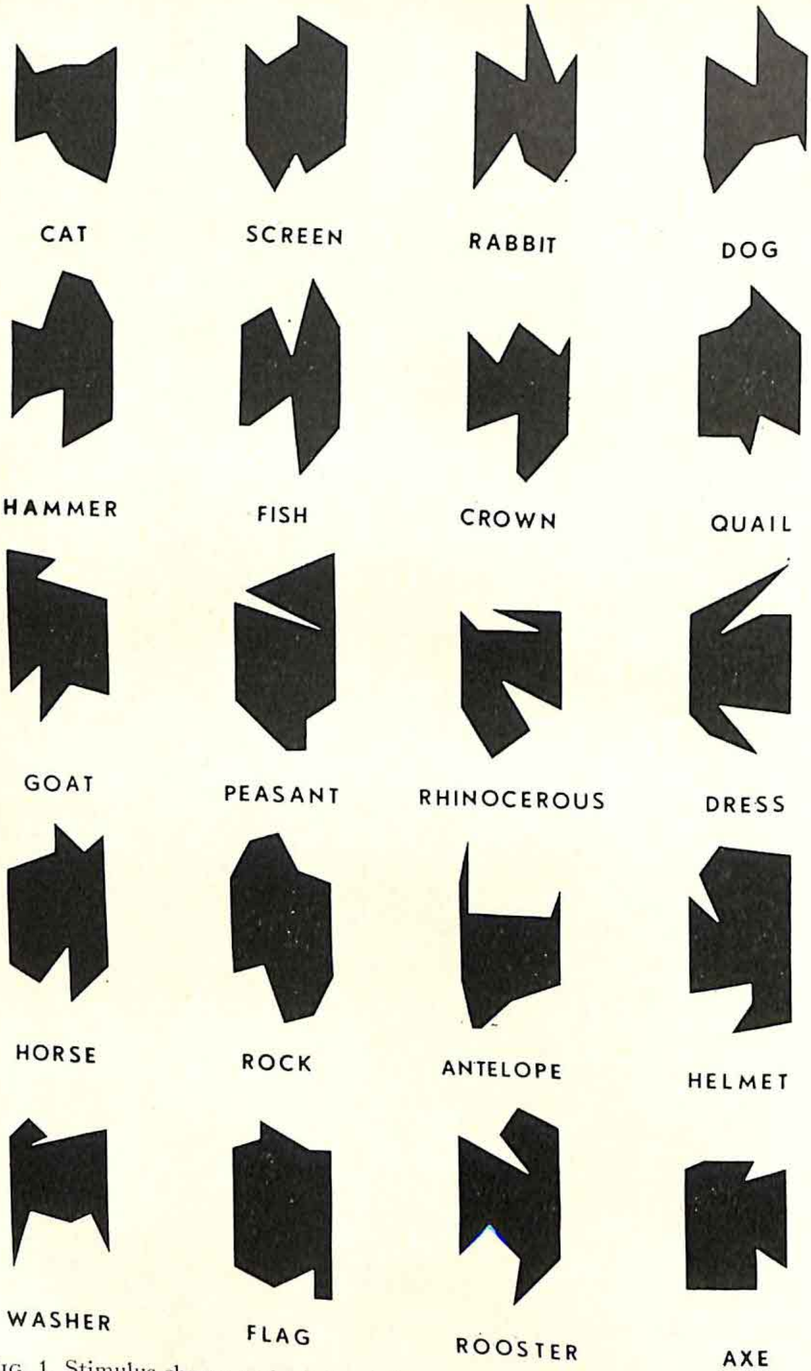


FIG. 1. Stimulus shapes and labels. (Top two rows were used for Exp. I–III.)

were randomly assigned to the trials. The test alternatives were constructed so that within a block of 32 trials the correct match occurred with equal frequency over the four positions of the alternative set. Furthermore, the alternatives were designed so that each of the eight shapes of the stimulus population occurred with equal frequency in all positions of the response set, and the frequency of occurrence of any pair of shapes in a set of alternatives was approximately equal. In other words, any shape occurred in an alternative set together with any given other shape about the same number of times (approximately 16) over the 64 recognition trials.

The test sequences for memory sets of Size 2 or 4 were constructed by using the same sequence of retention intervals and of four shape alternatives as used for a memory set of one, and by constraining the stimulus sequence so that the actual shape cued on a given trial was the same shape as a test item in the memory set of one. Therefore, the response balancing and sequence of testing stimuli was identical across memory set sizes of one, two, or four. Memory sets of Size 2 were constructed so that each stimulus shape occurred equally often in either position of the set. A given stimulus shape occurred in a memory set with four different shapes, and the sequence of cue lights was balanced over the two possible positions in blocks of 8 trials. Memory sets of Size 4 were constructed according to the same principles used in making up the four shape alternatives. Each shape occurred in each position equally often, and cue lights were balanced across the four possible positions in blocks of 32 trials.

### *Results and Discussion*

The mean number of incorrect choice responses in training was 18.05 for the Named condition and 19.12 for the Unnamed group, $t\ (58) = .45$, $p > .30$. The mean total naming errors in the Named condition was 13.63, and the mean naming errors on the last eight trials was .64.

Tables 1 and 2 show the mean percent correct choices for the recognition test as a function of naming condition, memory load, and retention interval. The main effects of name training, $F\ (1,\ 54) = 16.34$; memory load, $F\ (2,\ 54) = 25.76$; and retention interval, $F\ (7,\ 378) = 4.42$, all $p$s $< .01$, were statistically significant. Name training led to better performance in all conditions, and the advantage was relatively constant across retention intervals. Performance decreased for both Named and Unnamed conditions as memory load increased, and there was a relatively slight decline in performance as the retention interval lengthened. The retention interval did not interact with either naming condition or with memory load. The predicted interaction of Naming × Memory Load was highly significant, $F\ (2,\ 54) = 5.76$, $p < .01$. The interaction was still reliable at performance levels where the ceiling effect was less pervasive (two and four shapes), $F\ (1, 54) = 5.46$, $p < .05$.

TABLE 1

MEAN PERCENT CORRECT CHOICE RESPONSES IN EXPERIMENT I AS A FUNCTION OF NAMING CONDITION AND RETENTION INTERVAL

| Cond. | Retention interval (in sec.) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | .2 | 1 | 2 | 4 | 8 | 15 | 30 | 50 |
| Named | .92 | .93 | .91 | .95 | .89 | .89 | .89 | .86 |
| Unnamed | .80 | .78 | .79 | .85 | .78 | .86 | .74 | .74 |

TABLE 2

MEAN PERCENT CORRECT CHOICE RESPONSES IN THE TEST STAGE OF EXPERIMENT I AS A FUNCTION OF NAMING CONDITION AND MEMORY LOAD

| Cond. | Memory load | | |
|---|---|---|---|
| | 1 | 2 | 4 |
| Named | .97 | .90 | .84 |
| Unnamed | .95 | .82 | .60 |

These data partially support the conclusions drawn from the codability work (Brown & Lenneberg, 1954; Ervin-Tripp & Slobin, 1966). Verbal coding does appear to become more important as memory load is increased. However, the effects of name training did not interact with retention interval. The present data suggest that verbal encoding expands the amount of information that can be maintained but does not affect the temporal stability of the representation.

## EXPERIMENT II

The second experiment was designed to replicate the Naming × Memory Load interaction with a repeated-measures design and to obtain information concerning the underlying process. The results of the first experiment appear to be consistent with a rehearsal buffer notion of memory (Atkinson & Shiffrin, 1968). According to such a theory of memory, items can be maintained in short-term memory only by being recycled or rehearsed. It is quite possible that the constant advantage of the Named over the Unnamed condition in the present experimental paradigm can be explained as a rehearsal effect. The *S*s in the Unnamed condition might be able to maintain a reasonable image or visual representation of a limited amount of information. However, as the input load is increased, they are unable to rehearse the total load effectively, and thus they must attempt to keep only a portion of the stimulus elements. In other words, it is possible that *S*s in the Unnamed condition can maintain, or rehearse, a limited amount of information as an image. It appears likely that *S*s in the Named condition can effectively rehearse a much larger number of stimulus items by coding them verbally and rehearsing the verbal codes rather than attempting to maintain the visual representations. At the end of the retention interval, the Named *S*s could then decode the labels into the appropriate visual representations. This ability to rely on the verbal codes might expand the amount of information that can be rehearsed effectively. The hypothesis appears reasonable, but there has been little or no direct experimental evidence that verbal coding does in fact gain its advantage over iconic representation via rehearsal efficiency.

One means of exploring the rehearsal effects of verbal coding is to interfere with the rehearsal process. If the observed Naming × Memory Load interaction is due to more efficient rehearsal in the Named condition, then interfering with rehearsal during the retention interval should reduce or eliminate the interaction.

A relatively effective method of interfering with verbal rehearsal is to require *S*s to count backwards during the retention interval (Peterson & Peterson, 1959). The rehearsal hypothesis is tested in the second experiment by modifying the cued recognition task used in Exp. I to introduce an interference task of counting backwards during the retention interval on half of the trials. If the advantage of the Named condition arises from more efficient rehearsal, then the Naming × Memory Load interaction should be much stronger on trials without interference than on trials where interference is present.

### *Method*

*Design.*—The *S*s were given either Named or Unnamed training and then were observed in a recognition task with memory loads of two and four shapes. The *S*s were also given an interference task of counting backwards by one's on half of the recognition trials, resulting in a 2 × 2 × 2 design with repeated measures on both the interference and memory load factors. A total of 20 *S*s (introductory psychology students) served in the experiment with 10 *S*s in each of the training conditions.

*Procedure.*—The training task was the same as that used in the first experiment with eight randomly ordered runs through the set of eight stimulus shapes.

The testing procedure was similar to that used in Exp. I, but with several modifications. All *S*s were presented a memory set of two shapes on half of the 64 trials, and a memory set of four shapes on half of the trials. Presentation times for memory sets of two and four shapes were, respectively, 1 and 2 sec. At the end of an 8-sec. retention interval, *S*s were cued as in Exp. I to indicate the test position within the memory set. A set of four whole-shape alternatives was then presented for 4 sec., and *S*s were to respond in the same manner as in Exp. I before the 4-sec. intertrial interval terminated. On half of the recognition trials, *S*s were given a free (noninterference) retention interval, and on half of the trials *S*s were given an interference task during the retention interval. Trials without interference were conducted exactly as in Exp. I with a memory load of two or four shapes on any given trial. On interference trials, *S*s were given a three-digit number aurally as soon as the memory set went off the screen, and they were asked to count backwards by one's from this number, at a rate of one digit per second, during the 8-sec. retention interval, until a cue light came on.

Four trials within each block of eight were randomly assigned a memory load of two shapes, with the remaining four trials given a memory load of four shapes. Also, within each block of eight trials, half of the trials were assigned to the interference condition so that each block contained two interference trials with a memory load of two shapes and two interference trials with a memory load of four shapes. Cue lights were balanced over the 64 trials so that cues appeared with equal frequency over the possible positions within the memory set for both memory set sizes.

### *Results and Discussion*

The mean number of recognition errors in training was 13.3 for the Named and 13.5 for the Unnamed condition, $t$ (18) = .106, $p > .30$. The mean total naming errors in the Named condition was 13.0, with .4 the mean for the last eight trials.

Table 3 shows the mean percent correct in the test stage as a function of naming condition, memory load, and interference condition. The main effects of naming, $F$ (1, 18) = 16.21; memory load, $F$ (1, 18) = 62.2; and interference, $F$ (1, 18) = 126.99, all $p$s < .001, were all highly

TABLE 3

Mean Percent Correct in the Test Stage of Experiment II as a Function of Naming Condition, Memory Load, and Interference

| Naming Cond. | Memory load | |
|---|---|---|
| | 2 | 4 |
| Noninterference cond. | | |
| Named | .89 | .71 |
| Unnamed | .77 | .49 |
| Interference cond. | | |
| Named | .66 | .38 |
| Unnamed | .44 | .29 |

reliable. None of the two-factor interactions were statistically significant, including the predicted interaction of Naming × Memory Load. The three-factor interaction of Naming × Memory Load × Interference approached significance $F$ (1, 18) = 3.99, $p < .07$.

These results at best provide marginal support for the Naming × Memory Load interaction and the explanation of the effect in terms of a rehearsal effect of verbal coding. The data are in the predicted form in that naming seems to help more as the memory load increases, but only when rehearsal is allowed to take place. However, the results are not compelling, and it is possible that both ceiling and floor effects were operating and contributing to the trend toward a three-factor interaction. Without interference, the task was simply too easy with a memory load of two shapes and too difficult with interference and a memory load of four shapes.

Two further experiments were executed to examine the components of the possible three-factor interaction observed in Exp. II. In the interest of brevity, the experiments will be mentioned briefly. One of the experiments replicated the Naming × Memory Load interaction without interference, and the other replicated the same interaction with interference present. An attempt was made to avoid ceiling effects in the noninterference experiment by presenting the stimuli for half of the time used in the first two experiments. Floor effects were avoided in the interference experiment by using memory loads of one and two shapes. Both experiments were successful in achieving more reasonable performance ranges. Furthermore, the Naming × Memory Load interaction was present, $F$ (1, 28) = 7.05, $p < .05$, when *S*s did not have to count backwards, but was not reliable, $F$ (1, 30) = 3.14, $p > .05$, when *S*s were required to count backwards. Thus, these two experiments lend further credence to the marginal three-factor interpretation of the effects of verbal coding on recognition memory of shapes.

## Experiment III

The preceding experiments provide some support for the notion that verbal coding contributes to memory performance via an increase in rehearsal efficiency. Support for a rehearsal interpretation of the effects of verbal coding comes from the reduction of the Naming × Memory Load interaction when verbal interference is present. However, close examination of these data indicates several problems for a simple rehearsal interpretation. The first problem that might be raised is that even though the Naming × Memory Load interaction was reduced or eliminated by interference, the main effect of name training remained significant in all of the preceding experiments. Thus, it is possible that name training gains part of its advantage over iconic representation through a more efficient rehearsal scheme, but perhaps there is some other factor such as an acquired distinctiveness or stimulus differentiation which contributes to the general superiority of the name-training condition.

Another question that remains unanswered by the preceding experiments is to what extent is short-term-memory (STM) rehearsal possible in the Unnamed condition. Or phrased differently, can *S*s rehearse an image? The results of the first experiment might be interpreted to mean that *S*s can indeed rehearse an image since performance remained constant as a function of retention interval from .2 to 50 sec. However, it is possible that this result is obscured by a confounding of two different memory systems. Perhaps the constancy of performance as a function of memory interval in the Named condition was truly a result of an ability to rehearse the memory set, but in the Unnamed condition the same constancy might have resulted from either rehearsal in STM or from a reliance on a relatively permanent long-term-memory store (LTM).

Consideration of recent theoretical accounts of general memory processes suggests a task which permits an investigation of rehearsal effects in STM. Atkinson & Shiffrin (1968) have developed a general model of memory which incorporates a rehearsal buffer and a transfer process from the buffer to long term memory. According to Atkinson and Shiffrin's model, the strength of an LTM trace is directly determined by the total amount of time that an item spends in the rehearsal buffer. Information about an item is transferred into LTM as long as the item stays in the short-term rehearsal buffer.

This model suggests a reasonable way to determine the importance of rehearsal for both Named and Unnamed conditions in the present series of experiments. In accord with the model, let us assume that the strength of the LTM trace is a direct function of the amount of time spent in STM. It is also reasonable to assume that, at least in the Named condition, *S*s will rehearse the verbal codes for as much time as they are allowed (within reason) and that this rehearsal will contribute to the strength of the LTM trace of the labels, and possibly the visual representation, if it exists separately.

These assumptions imply an increase in probability correct as a function of retention interval (i.e., rehearsal time). The implication holds only if *S*s respond on the basis of LTM. Therefore, the third experiment attempts to monitor the strength of LTM, while assuming this strength to be a direct function of time spent in the rehearsal buffer. One way to perform the experiment is to allow items to remain in STM for varying amounts of time, and then to interfere with the rehearsal in an extremely disruptive manner, then give a recognition task. One very disruptive interference task is to require *S*s to count backwards by three's in time with a metronome during the retention interval (Peterson & Peterson, 1959). It should be noted that the interference task used in Exp. II was probably considerably less disruptive than counting backwards by three's.

The task paradigm for Exp. III is as follows: stimulus presentation (1 sec.), preinterference interval (2 or 7 sec.), interference interval (counting backwards by three's for 16 sec.), and recognition test. This paradigm leads to several alternative predictions which depend on the rehearsal assumptions for the Unnamed condition. In general, if rehearsal is effective in transferring information to LTM, performance should improve as a function of the preinterference interval. On this assumption, performance in the Named condition can be expected to improve as the preinterference interval is lengthened. Performance for the Unnamed condition will improve at the same rate if their representations can be rehearsed and transferred at the same rate as the verbal codes of the Named condition. The improvement in performance as a function of preinterference interval should be less for the Unnamed condition than the Named to the extent that iconic rehearsal is less efficient than verbal rehearsal in maintaining information and transferring it to LTM

### *Method*

*Design.*—Two types of training, Named and Unnamed, and a recognition task with 2 preinterference intervals (2 or 7 sec.) were used in a 2 × 2 design with repeated measures on the preinterference interval factor. A total of 32 *S*s (introductory psychology students) participated in the experiment with 16 *S*s serving in each of the name-training conditions.

*Procedure.*—Training for both Named and Unnamed *S*s was identical to that used in the preceding experiments.

The recognition test was basically the same as those used in previous experiments with a memory load of two shapes. On every trial, *S* was presented with a memory set of two shapes for 1 sec. The *S* was to remember the memory set for a total retention interval of 18 or 23 sec., after which he received a cue light designating the test position in the memory set. A set of four whole-shape alternatives was then presented for 4 sec., and *S* made his choice from this set. The retention interval on each trial was divided into two parts. The first part consisted of a preinterference interval of either 2 or 7 sec. At the end of this preinterference interval, an electronic metronome began beating at a rate of one beat every 2 sec. for 16 sec. The *S* was given a three-digit number at the onset of the metronome, and he was instructed to count backwards by three's in time with the metronome during this 16-sec. interference period. The interference period was followed immediately by a cue light and then by the set of whole-shape alternatives.

The 2- and 7-sec. preinterference intervals were assigned randomly in 8-trial blocks to half of the 32 recognition trials. Cue lights and position of correct response within the alternative set were balanced in blocks of 16 trials.

### *Results and Discussion*

The mean number of incorrect choice responses in training was 18.4 for the Named condition and 18.7 for the Unnamed condition, $t$ (30) = .152, $p > .30$. The mean total naming errors was 13.9, and the mean naming errors on the last eight trials was .53.

Table 4 shows the mean percent correct for the recognition test as a function of name training and preinterference interval. Although the level of performance is reasonable, with no apparent ceiling or floor effects, these data exhibit a rather striking lack of differences. There were no significant main effects or interactions present. Recognition performance remained unchanged as a function of preinterference interval length, and for the first time there was no main effect of name training.

The possibility arises that in recognition memory, linguistic coding has primarily STM effects. Verbal coding may influence STM retention by providing an efficient rehearsal scheme. This rehearsal advantage may be more important as the number of shapes to be remembered increases. However, if the recognition task demands a heavy reliance on LTM, the verbal code may be mostly irrelevant. This interpretation would be compatible with a dual storage format in LTM. With such a system, an iconic representation would be maintained distinct from a verbal code. The verbal code would be closely associated with the iconic representation and, to the extent that the code specified particular features of the stimulus, it might provide an important source of redundancy in recognition. However, with complex shapes and the general labels that have been used in these experiments, the redundancy effects are apparently minimal, as evidenced by the outcome of Exp. III. To the extent that *S*s relied upon the names, performance should have increased as a function of preinterference interval (Peterson & Peterson, 1959, Exp. II).

TABLE 4

Mean Percent Correct in the Test Stage of Experiment III as a Function of Naming Condition and Preinterference Interval

| Naming Cond. | Preinterference interval (in sec.) | |
|---|---|---|
| | 2 | 7 |
| Named | .65 | .63 |
| Unnamed | .63 | .63 |

Exp. III, when taken in conjunction with the previous experiments, provides the strongest support for a hypothesis which limits the effect of verbal coding in recognition to the short-term rehearsal buffer. Exp. III is the first experiment where it is reasonable to assume that the rehearsal buffer is not playing an important role (due to the stringent interference used), and this is the only experiment in the series where the absolute effect of name training is absent.

## Experiment IV

From the results of Exp. III, the hypothesis has been offered that the effects of verbal coding on recognition arise from rehearsal efficiency and are therefore limited to STM. This hypothesis implies that the effects of naming will initially be present, but then should disappear in a task extending beyond the short-term span.

One such task is a continuous "old–new" recognition task, where *S*s are presented a sequence of items and are asked to judge whether each item is old or new. The first time an item appears it is considered to be new. The occurrence of an old item can vary with respect to lag, i.e., the number of items that have intervened since it was first seen as a new item. Such a task offers a range of test intervals or lags including those lags that are clearly within the STM span to those that are clearly beyond the short-term span. On the rehearsal hypothesis, the effects of verbal coding on recognition performance should be restricted to lags within the range of STM.

However, pilot data revealed that a normal old–new recognition procedure is not sufficiently difficult to preclude ceiling effects. Therefore, another measure of the rehearsal hypothesis was necessary. One possible measure is an estimate of recency in which *S* must indicate how many shapes have intervened since he saw a given shape. In other words, on an item that *S* considers old, he is to attempt to estimate how old it is, i.e., to guess the actual lag since its last presentation. In general, recency estimates might be based upon somewhat different information than a simple recognition response. However, in the present situation, *S*s would have been exposed to the shapes several times during training, and so even a simple old–new recognition task relies upon a type of recency discrimination.

The predictions of a rehearsal model of verbal coding are basically the same with a recency task as for the simple old–new task. If a verbal code simply increases rehearsal efficiency in STM, then estimates of recency should be more accurate for the Named condition, but only for lags within the short-term retention span.

## *Method*

*Stimuli.*—The stimulus shapes (Fig. 1) were 10-sided nonsense shapes drawn from a pool similar to that used for the shapes of the preceding experiments. Twelve shapes were added to the set used in the first five experiments to make a total of 20 shapes. The shapes were arbitrarily assigned names as in the previous experiments.

*Design.*—Two training conditions, Named and Unnamed, and a recognition task with 2 trials, 2 different orders of presentation, and 5 lag lengths (1, 3, 4, 5, and 10) were used in a $2 \times 2 \times 2 \times 5$ design with repeated measures on all factors except type of training. A total of 28 *S*s (undergraduate psychology students) participated in the experiment with 14 *S*s assigned to each of the 2 training groups.

*Procedure.*—The *S*s received a modified version of the training procedure used in previous experiments. As before, on each trial, *S* was shown a whole shape for 2 sec. and then was asked to give the name for the shape when *S* was assigned to the Named condition. After a 3-sec. delay, *S* was shown a set of two half-shape alternatives and was asked to choose the half shape that matched the stimulus shape he had seen 3 sec. earlier. The task was identical to that used in previous experiments except that the alternative set consisted of 2, rather than 4, half shapes. Training continued for a total of eight randomly ordered runs through the set of 20 shapes. The sequence was designed so that each shape was tested four times against a set of bottom contours. The 50% occurrence of top and bottom contours was randomized in blocks of 20 trials, as was the correct response position within the alternative set.

In the test task, *S* was presented with a sequence of 40 stimulus shapes. Each trial consisted of presenting a single shape for 1 sec., with an intertrial interval of 6 sec., followed by a warning light for 1 sec. Every time a shape was presented, *S* was to make a verbal indication as to whether the shape was new or old in that sequence. There were only 20 separate shapes, each of which *S* had seen several times during training. Therefore, *S* had seen all of the shapes before, but he was to say "new" the first time a particular shape appeared in the 40-trial sequence, and he was to indicate that the shape was old the second time he saw it in the sequence. On those trials that *S* thought the shape was new, he was to say "new" aloud. If *S* thought the shape was old, he was to give a number to indicate how many other stimuli had occurred since he last saw the shape (i.e., his estimate of the actual lag). The number was to reflect *S*'s best estimate of the shape's recency or the number of intervening stimuli. The *S* was presented one sequence of 40 shapes and, after a delay of approximately 3 min., he was given a second sequence consisting of a different ordering of the same set of 20 shapes.

Test sequences were constructed by constraining the 40 positions in the sequence so that each shape appeared twice, once as new and once as old. Furthermore, only five lag lengths were allowed (1, 3, 4, 5, and 10), with each lag length appearing four times. A lag length of one meant that one shape intervened between the first and second occurrence of a particular shape. The other lags were defined similarly. One test sequence was constructed, and a forward and backward ordering of this sequence was used for the two recognition test sequences.

## *Results and Discussion*

The mean number of recognition errors in training was 22.72 for the Unnamed condition and 23.30 for the Named condition, $t$ $(26) = .13$, $p > .30$. The mean

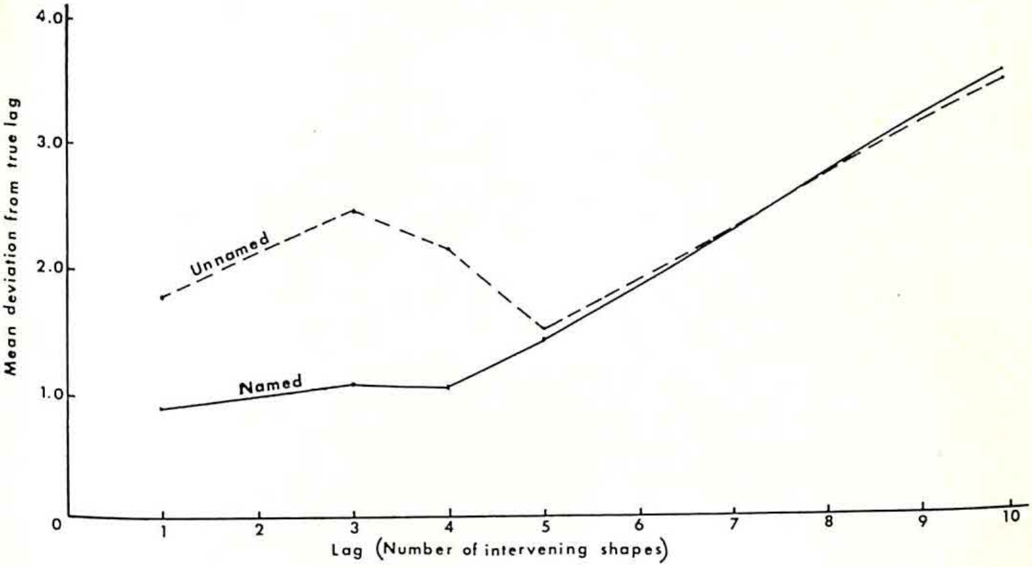


FIG. 2. Recency Estimates for Exp. IV as a function of naming condition and lag (number of intervening shapes).

total naming errors in the Named condition was 39.61, and 2.86 for the last 20 trials.

The recency judgments as a function of name training are presented in Fig. 2. The mean of each *S*'s unsigned deviations from the true lag was used in the analysis. There were significant main effects of name training, $F\ (1,\ 26) = 39.10$, $p < .001$, and lag $F(4,\ 104) = 36.92$, $p < .001$. The Named condition was more accurate than the Unnamed condition when the data are collapsed over all other factors. The significant main effect of lag is primarily due to all *S*s performing worse on the lag of 10. Performance was relatively constant over the middle range of lags. Main effects of sequence, $F\ (1,\ 26) = 2.80$, $p > .10$, and trial, $F\ (1,\ 26) = .20$, $p > .10$ were not significant.

The predicted Name Training × Lag interaction was reliable, $F\ (4,\ 104) = 4.58$, $p < .01$, with the Named condition more accurate on lags of Length 4 or less. As the lag increased beyond 4, the difference between the training conditions disappeared. The observed interaction is in complete accord with the rehearsal model of the effects of verbal coding on recognition memory. Name training leads to more accurate estimates only in the STM range. This effect is very likely due to the Named *S*'s ability to rehearse effectively a larger number of items than is possible in the Unnamed condition. The order of stimulus presentation (sequence) did interact with Lag, $F\ (4,\ 104) = 4.06$, $p < .01$. This interaction is due primarily to a relatively poorer performance on Sequence 2 at a lag of 10. In Sequence 2, lag lengths of 10 tended to appear later in the list than they appeared in Sequence 1, which could explain the difference in performance on lags of 10.

None of the three-factor interactions were significant, indicating that the observed Naming × Lag interaction generalizes over the range of both trials and sequences used in this experiment. The four-factor interaction of Naming × Sequence × Lag × Trial was statistically significant, $F\ (4,\ 104) = 3.28$, $p < .05$. There is no obvious interpretation of this interaction.

## GENERAL DISCUSSION

This series of experiments has partially clarified the effects of verbal coding in recognition memory and has made some progress in identifying the mechanisms of these effects. The results of the first experiment suggest that labeling increases the capacity of the short-term memory buffer, but has little effect on the temporal stability of the representation. Experiment II succeeded in replicating the interaction of name training with memory load and provides some indication that verbal coding gains its relative advantage through an efficient rehearsal mechanism. The last two experiments provide even stronger support for a rehearsal interpretation of verbal coding.

In Exp. III, a task with strong interference eliminated all effects of name training. The experiment gave strong support to the notion of a rehearsal effect of name training. Furthermore, the results of Exp. III suggest that *S*s do not rely on a verbal code for shape recognition in LTM. The recency estimates in the last experiment support this hypothesis.

These experiments leave little doubt that linguistic coding can affect recognition memory of visual stimuli. The verbal code appears to be particularly important in STM tasks where it leads to more efficient rehearsal and an expanded STM capacity. The recency data from Exp. IV do not completely exclude the possibility that a verbal code can be used in a simpler recognition from LTM. However, these data in conjunction with the data from Exp. III suggest that there is little or no effect of verbal coding in recognition from LTM. The *S*s seem to rely on the visual representation without reference to the verbal label when forced to respond on the basis of LTM. This strategy is undoubtedly dependent upon the quality of the label. In these experiments, the labels were relatively arbitrary and so did not contain much information about the shape. It is quite possible that a more "relevant" verbal code can provide redundant information about the visual image.

Finally, it must be noted that there has been no attempt to confine all of the effects of linguistic coding to STM. The rehearsal advantage in STM appears dominant in recognition tasks. However, it is quite likely that verbal coding is important in other task paradigms such as free recall. The verbal code provides access to a rich and flexible retrieval system. One might expect strong effects of verbal coding in tasks where retrieval from LTM is an important component.

## REFERENCES

ARNOULT, M. D. Recognition of shapes following paired associates pretraining. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research.* (Publication No. 455) Washington, D. C.: National Academy of Science, National Research Council, 1956.

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system. In Spence & Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

BROWN, R. W., & LENNEBERG, E. H. A study in language and cognition. *Journal of Abnormal and Social Psychology*, 1954, **49**, 454–462.

ELLIS, H. C. *The transfer of learning.* New York: MacMillan, 1965.

ELLIS, H. C., BESSEMER, D. W., DEVINE, J. V., & TRAFTON, C. L. Recognition of random tactual shapes following predifferentiation training. *Perceptual and Motor Skills*, 1962, **14**, 99–102.

ELLIS, H. C., & MULLER, D. Transfer in perceptual learning following stimulus predifferentiation. *Journal of Experimental Psychology*, 1964, **68**, 388–395.

ERVIN-TRIPP, S. M., & SLOBIN, D. I. Psycholinguistics. *Annual Review of Psychology*, 1966, **17**, 435–474.

GLANZER, M., & CLARK, W. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 289–299.

GLANZER, M., & CLARK, W. The verbal loop hypothesis: Binary numbers. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 301–309.

HAGEN, J. W., & KINGSLEY, P R. Labelling effects in short term memory. *Child Development*, 1968, **39**, 113–121.

HAKE, H. W., & ERICKSEN, C. W. Role of response variables in recognition and identification of complex visual forms. *Journal of Experimental Psychology*, 1956, **52**, 235–243.

JEFFREY, W. E., & BOGARTZ, R. S. The criterion task as a factor in positive transfer following acquired distinctiveness pretraining. *Psychological Reports*, 1961, **8**, 15–19.

LANTZ, D., & STEFFLRE, V. Language and cognition revisited. *Journal of Abnormal and Social Psychology*, 1964, **69**, 472–481.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

PETERSON, L. R., & PETERSON, M. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

RANKEN, H. B. Effects of name learning on serial learning, position learning, and recognition learning with random shapes. *Psychological Reports*, 1963, **13**, 663–668.

ROBINSON, J. S. The effect of learning labels for stimuli on their later discrimination. *Journal of Experimental Psychology*, 1955, **49**, 112–114.

SANTA, J. L., & RANKEN, H. B. Verbal coding and redintegrative memory for shapes facilitated by naming. *Psychonomic Science*, 1968, **13**, 109–110.

SPIKER, C. C. Stimulus pretraining and subsequent performance in the delayed reaction experiment. *Journal of Experimental Psychology*, 1956, **52**, 106–111.

(Received July 19, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 279–282

# ASSIMILATION AND CONTRAST EFFECTS IN VISUAL DISCRIMINATION BY RHESUS MONKEYS[1]

MARTHA WILSON[2]

*University of Connecticut*

Sets of visual stimuli which differed on one of three dimensions, length, orientation, or density, were judged by 12 rhesus monkeys in a paired-comparison procedure. It was found that the discriminability of any pair of stimuli was a function of two factors: the physical distance between the stimuli and the relationship of the stimuli to the adaptation level of the series. The results point to categorical perception in a nonspeech mode, independent of identification training.

A great deal of evidence suggests that a special perceptual decoder is required for the analysis of speech as opposed to nonspeech stimuli. The existence of categorical perception has been taken as one indicator that a unique mechanism is involved in the encoding and decoding of speech stimuli (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, for a comprehensive review). Several recent experiments have examined the question of whether categorical perception of nonspeech stimuli can be demonstrated, and this problem has been taken as critical for disconfirming the theory referred to above. Although the experimental conditions were almost identical, two studies have provided contradictory results. It appeared in the Cross, Lane, and Sheppard study (1965, Exp. II) that discrimination of stimuli in two different categories was better than that within either category. A later study, which attempted to reproduce this result, failed to find superior intercategory as opposed to intracategory discrimination (Parks, Wall, & Bastian, 1969). Both studies used verbal labels to define the two categories.

The results presented here arose from entirely different considerations. Sets of visual stimuli on three dimensions were judged by rhesus monkeys so that perceptually equivalent, unidimensional stimulus sets could be constructed. However, the results also provided evidence for categorical perception in a nonhuman primate in a nonspeech mode.

## METHOD

*Subjects.*—Twelve sophisticated rhesus monkeys, previously trained on visual object and pattern discrimination, were used as *S*s. Of the original sample of 14, 2 failed to learn the initial phases and were discarded.

*Apparatus.*—Testing was done in a Wisconsin General Test Apparatus (WGTA) with a stimulus board having foodwells spaced 32 cm. apart. Stimulus sets for each dimension consisted of 10 white cardboard squares, 6.5 cm. on a side, with an appropriate figure centered on each card. These were covered with transparent plastic. Two identical sets were constructed and were interchanged frequently during testing. All stimulus figures were bars, 4 mm. wide and were outlined in black Chartpak tape, $\frac{1}{32}$ in. wide. Stimuli were constructed from Letratone printed papers. The two dimensions not varied in a given set were made constant at the middle values of these dimensions. Stimulus values are given in Table 1, and representative stimuli are shown in Fig. 1.

*Procedure.*—Monkeys were given 30 trials/day with a raisin reward for choosing the correct stimulus. First, *S*s were trained to choose between Stimulus 1 (+) and Stimulus 10 (−) on a given dimension to a criterion of 28 correct in 30 trials.[3] Then all 10 stimuli were presented in a paired comparison design. The stimulus in a pair that was closer to the positive stimulus in training was rewarded. Each pair was presented two times, making a total of 90 judgments by each monkey on

---

[1] This research was supported by Grant MH-18217 from the National Institute of Mental Health, United States Public Health Service. The author is grateful to Peter Van Gelder for his helpful comments.

[2] Requests for reprints should be sent to Martha Wilson, Psychology Research Building, U-152, University of Connecticut, Storrs, Connecticut 06268.

[3] The reward contingencies were inadvertantly reversed on the density dimension, so that Stimulus 1 was avoided rather than approached.

TABLE 1

Physical Values of Stimuli in Three Dimensions

| Dimension | Stimulus value | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Length (cm.) | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 | 5.5 |
| Orientation (degrees) | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| Density (lines/cm) | 3.0 | 4.0 | 4.8 | 6.2 | 9.5 | 12.5 | 15.8 | 19 | 22 | 23 |

each dimension. Every stimulus appeared equally often, once in each position for a given pairing. A quasirandom order of presentation insured that no particular stimulus value was presented on 2 successive trials, and neither position was rewarded on more than 3 successive trials. All monkeys judged the three dimensions in the same order: length, orientation, density.

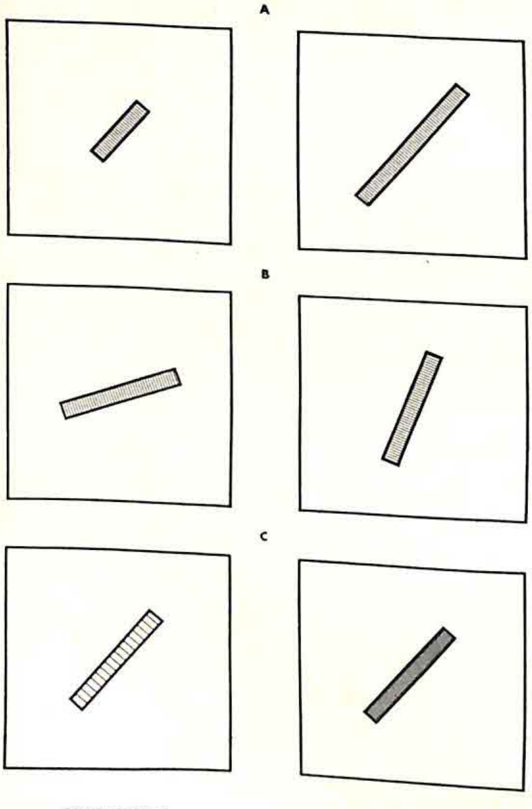


Fig. 1. Representative stimuli on (A) length, (B) orientation, and (C) density dimensions. (The two dimensions not varied in a given stimulus set were kept constant at the middle value of those dimensions.)

## Results

The proportion of choices of the rewarded stimulus in each pair was calculated for all monkeys and the scale value of each stimulus in the set, based on Thurstone's Case V, was derived as shown in Guilford (1954). The scale values obtained are shown in Fig. 2. The algebraic signs of the scale values are retained since the scale value of 0 represents a true psychological zero (cf. Guilford, 1954, p. 171). The stimulus which is given a scale value of 0 thus represents the adaptation level (AL) or neutral point of the series (Helson, 1964).

It was noted during testing that responses to pairs in which the stimuli were closely spaced seemed to be correct more often when the stimuli were somewhere in the middle of the range than when they were at the extremes. Since a first approximation to the AL for a stimulus series without an anchor is also somewhere

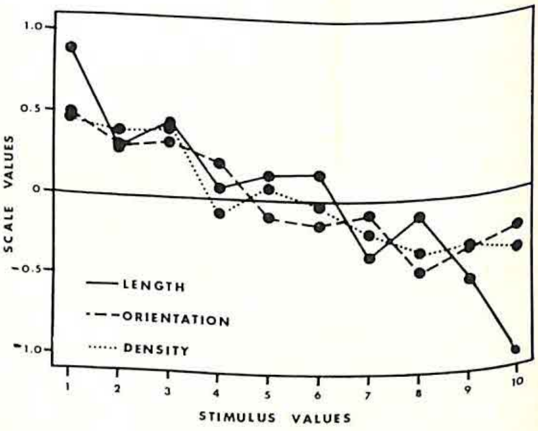


Fig. 2. Scale values of 10 stimuli in each dimension derived from paired comparisons.

in the middle of the stimulus range, it seemed worthwhile to explore the possibility that this phenomenon depended upon the relationship of the stimuli to the AL. This led to an analysis of performance in terms of two factors: the number of steps between the stimuli in a pair, and the relationship of the stimuli to each other in terms of the AL for the series.

The neutral point for each dimension is, as noted above, that stimulus value which yields a scale value of 0 on the stimulus versus scale value curves. Stimulus pairs were classified into those in which both stimuli were on the same side of the AL (same algebraic sign) and those in which the stimuli were on different sides of the AL (opposite algebraic signs). The frequency of correct choices was then computed for each number of steps of separation between the stimuli. For example, two-step separations of stimuli lying on the same side of the AL for length (Stimulus 1 vs. 3; 2 vs. 4; 3 vs. 5; 7 vs. 9; etc.) were compared with two-step separations of stimuli lying on either side of the AL (5 vs. 7; 6 vs. 8). This was done for all step separations from one to five: these constituted all of the possible stimulus pairs in which a given step difference could lie either on one side or both sides of the AL.

The results of the analyses are shown in Fig. 3. In all three dimensions, with the exception of one unexplained value for length, and one for density, there is a striking improvement in discriminability of stimulus differences when the stimuli lie on opposite sides of the AL. For example, a stimulus difference of 20° in orientation on one side of AL was not discriminable ($p = .53$), while the same difference was clearly discriminated ($p = .65$) when the stimuli lay on opposite sides. There is a tendency for the differences in discriminability to become smaller as the steps get larger. However, if only the smaller stimulus steps are considered, the results appear to approach categorical perception. The overall difference between the two kinds of stimulus pairs is significant for length ($\chi^2 = 13.26$, $p < .001$), orientation ($\chi^2 = 20.83$, $p < .001$), and density ($\chi^2 = 10.40$, $p < .01$). The better discrimination noted in the middle of the range is thus seen to depend upon the way in which the AL structures the dimensional perceptually. Furthermore, equal sense distances derived from average discriminability may mask the differential sensitivity to stimulus differences that exists in different parts of the stimulus range.

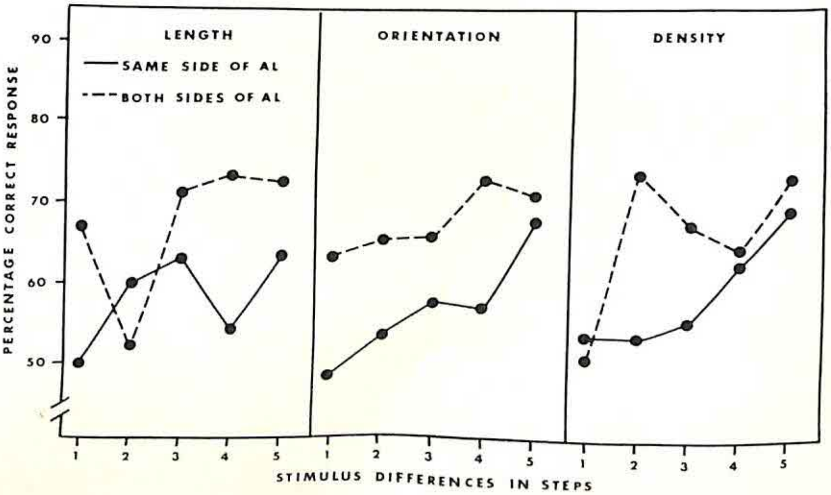


FIG. 3. Percentage of correct choices as a function of the physical differences between the stimuli, and the relationship of the stimuli to the adaptation level.

## DISCUSSION

These results show that the perceptual characteristics of stimuli depend upon their relationship to the prevailing AL for stimuli in that universe. It has previously been shown that the *identifiability* of a visual stimulus depends upon its psychophysical distance from AL (Wilson, 1971): these data show that the *discriminability* of stimuli also depends upon the way in which the AL structures the stimulus set.

The relevance of these data for the question of categorical perception in a nonspeech mode is obvious. The neutral point divides the stimulus set into two categories, longer/shorter, denser/less dense, etc., and discrimination within categories is more difficult than between categories. These effects have been studied as contrast and assimilation effects, and their dependence upon a single underlying mechanism has been clearly spelled out (Helson, 1964).

While the AL can be changed by variations in stimulus input, the absence of an explicit anchor in the Cross et al. (1965) experiment makes it likely that their stimulus continuum was divided in the middle of the stimulus range. This could explain the differences in discriminability obtained without the necessity of invoking the identification training. On the other hand, since there were only four stimuli in the set, the two extreme stimuli should be highly identifiable, while the two middle stimuli should be highly discriminable. This would lead to the result obtained by Parks et al. (1969). Only with a larger number of stimuli, as in the present experiment, should differences in intra- and intercategory judgment be obvious.

These data provide evidence for tendencies toward categorical responding in a nonhuman primate which are totally independent of identification training. The effects can be explained in terms of AL theory, which postulates a labile, stimulus-sensitive, interval level to which stimuli are referred. This kind of mechanism appears superficially to be of a different order than that responsible for the invariances characteristic of speech decoding. However, one could speculate that speech perception is categorical because the discontinuous articulatory mechanisms underlying speech production provide extreme and stable anchors to which speech stimuli are assimilated. On this view, speech production and perception are intimately associated, as the motor theory suggests, but via a perceptual mechanism that is shared with nonspeech perception.

## REFERENCES

CROSS, D. V., LANE, H. L., & SHEPPARD, W. C. Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. *Journal of Experimental Psychology*, 1965, **70**, 63–74.

GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

HELSON, H. *Adaptation level theory.* New York: Harper & Row, 1964.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431–461.

PARKS, T., WALL, C., & BASTIAN, J. Intercategory and intracategory discrimination for one visual continuum: Contributions of identification training and of individual differences. *Journal of Experimental Psychology*, 1969, **81**, 241–245.

WILSON, M. Shifts in categorization and identifiability of visual stimuli by rhesus monkeys. *Perception & Psychophysics*, 1971, **10**, 271–272.

(Received July 23, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 283–289

# INDEPENDENCE BETWEEN BINOCULAR RIVALRY SUPPRESSION DURATION AND MAGNITUDE OF SUPPRESSION [1]

ROBERT FOX [2] AND RONALD CHECK

*Vanderbilt University*

The elevation in threshold for test-probe stimuli presented to the eye during phases of suppression reveals an inhibitory effect of binocular rivalry (the suppression effect). Heretofore, tests of the suppression effect have been limited to stimuli presented at the start of suppression. This experiment examined the relationship between magnitude of the suppression effect and duration of suppression phases by presenting probe stimuli at representative temporal intervals throughout suppression phases. Using three trained *S*s, forced-choice form recognition thresholds were obtained for three temporal points after onset of suppression. The results of all *S*s indicate that recognition thresholds are significantly raised by an equal amount at all temporal points. These data do not support rivalry models that assume correlation between magnitude of suppression and suppression duration, and they suggest that magnitude and duration are controlled by separate mechanisms.

The cardinal characteristic of binocular rivalry is the phenomenal suppression of potent stimuli for durations on the order of seconds. It would be surprising if the powerful inhibitory process implied by suppression were confined to the phenomenal domain and did not exert more general effects upon vision. Evidence in favor of more general inhibitory effects has recently been obtained using the test-probe paradigm, whereby the relative fate of a test stimulus presented to an eye during phases of suppression and phases of dominance provides an indirect measure of the effect of suppression. The data from such experiments (e.g., Fox & Check, 1966, 1968; Wales & Fox, 1970) reveal that the energic threshold of the visual system is raised during suppression on the order of .5 log unit. This elevation, which represents a true reduction in sensitivity, is referred to as the suppression effect.

It is not known whether the suppression effect endures throughout the duration of the phase of suppression since test probes have always been presented at the start of phases of suppression and dominance. The purpose of the present experiment was to obtain direct evidence on that point by presenting test probes at various temporal points within suppression and dominance phases.

There are three hypotheses concerning the relationship between phase duration and the magnitude of the suppression effect that can be reasonably considered. To facilitate discussion of these hypotheses, it is helpful to conceptualize suppression in terms of a hypothetical inhibitory pulse that acts to raise the threshold of the afferent channel by some amount and thereby induces phenomenal suppression. The pulse may be conceptualized as a single pulse of several seconds duration or as the envelope of many pulses or brief duration. The magnitude of suppression corresponds to the height of the pulse and the duration of suppression corresponds to the width of the pulse.

The first hypothesis, which will be referred to as the "correlative hypothesis," involves the assumption that variations in the duration of suppression phases reflect variations in the magnitude of the suppression extant during a particular suppression phase. In the first version of this hypothesis, whose wave form is illustrated in Fig. 1a, it is assumed that the magnitude of suppression is maximum at the beginning of the suppression phase and then declines throughout the phase, reaching a level

---

[1] This research was supported in part by United States Public Health Service Grant EY 00590 to the senior author.

[2] Requests for reprints should be sent to Robert Fox, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203.

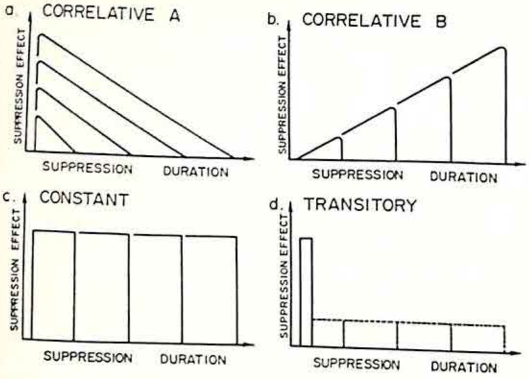


FIG. 1. Hypothetical inhibitory pulse waveforms based on the four hypotheses. (Wave Forms *a* and *b* represent two versions of the correlative hypothesis; Wave Form *c*, the constant hypothesis; Wave Form *d*, the transitory hypothesis. The dotted lines represent the assumption that phenomenal suppression continues after inhibition has become negligible).

where suppression is absent and the phenomenal state changes to dominance. Variations in the initial magnitude of suppression would produce concomitant variations in the time required for its decay, and these variations in decay would correspond to variations in the duration of suppression phases. This kind of wave form would seem to be most compatible with a model of rivalry based upon reciprocal inhibition that would make assumptions about fatigue or adaptation processes to account for the shifts of phenomenal state between the eyes.

A second version of the correlative hypothesis, illustrated in Fig. 1b, represents the assumption that the magnitude of suppression increases proportionally with the duration of the inhibitory pulse. On this wave form, increases in the peak magnitude of suppression correspond to increases in suppression duration. Both Wave Forms *a* and *b* are parsimonious in that they account for suppression magnitude and duration by one mechanism, yet each yields a different prediction about the magnitude of suppression that would be found at various temporal points within a suppression phase.

An alternative to the correlative hypotheses is the "constant hypothesis," which involves the assumption that the duration of a suppression phase and the magnitude of the suppression effect are independent quantities (see Fig. 1c). The prediction derived from this wave form is that the magnitude of suppression is a constant value independent of the point of onset of the test probe within the suppression phase.

Finally, there is the "transitory hypothesis," which makes the assumption that the suppression effect is present only during the beginning of a suppression phase, with the consequence that suppression effects could be obtained only at the commencement of suppression. An attempt to depict this wave form is shown in Fig. 1d. However, the wave form concept is not so helpful here because no causal relationship is posited; instead, the assumption is made that the inhibition responsible for the suppression effect is present only at the start of suppression. The derivative prediction is that the suppression effect would not be found for test probes presented relatively late within the suppression phase. The transitory hypothesis seems most similar to the concept of temporal inertial effects in the balance model of rivalry described by Levelt (1965).

The constant hypothesis and the transitory hypothesis are both less parsimonious than either version of the correlative hypothesis in that the constant hypothesis requires a separate mechanism for terminating suppression and the transitory hypothesis requires one or more mechanisms for maintaining and terminating suppression.

## METHOD

*Subjects.*—Two males, DC and CM, and one female, KK, ranging in age from 21 to 25 yr., served as paid volunteers. All *S*s were highly practiced in reporting rivalry and experienced in psychophysical observations. The *S*s DC and KK were naive regarding the purpose of the experiment, while CM was not.

The vision of each *S* was measured by modified Orthorater tests. The *S* DC had excellent uncorrected vision, Snellen acuity 20/10, and was orthophoric, while CM and KK had corrected-to-normal vision, Snellen acuity 20/25 or better, and scored less than .5 prism diopter on tests of phoria.

*Apparatus.*—The basic optical system was a large Wheatstone-type stereoscope constructed from optic bench components. The left-eye and right-eye stimulus fields were 6° squares located 63 cm. from each eye. An accessory optic channel in each field permitted optical superimposition of test stimuli on the plane of the rivalry targets. The entire apparatus, including light baffles and septum, was enclosed in a large light-tight housing. A set of trial frames was permanently mounted at the viewing port; these were fitted with 2-mm. artificial pupils with an extruded conical section that located the physical aperture as close as possible to the cornea and with refractive lenses for CM and KK. Either a chin cup and headrest or a biteboard assembly was also located at the viewing port. The stimulus holders and other optical components of the stereoscope were readily adjustable so that fusion could easily be obtained and held by each *S* without strain.

The configuration of the rivalry targets is illustrated in Fig. 2. The target with the black center disk was presented to the left eye for CM and KK and to the right eye for DC. Beyond the outer edges, the surround of each target was black. The peripheral concentric rings served to maintain fusion of the targets. When the targets were fused, the center disks were coincident, and it was the rivalry of these targets that *S*s reported.

The targets were constructed from high contrast negatives (Kodalith) and mounted on translucent Plexiglas sheets, a combination which yielded homogeneous white areas within the clear portions of the negatives and essentially opaque areas for the black portions of the negatives. In the stereoscope apertures, the Plexiglas sheets were transilluminated by incandescent sources. The luminance level of the targets, as measured at the viewing port, was varied for each *S* by means of neutral density filters: For CM, the luminance levels for the white areas of each target were .32 (right eye) and .38 ftl. (left eye); for DC these values were .43 and .29 ftl.; for KK these values were 1.36 and 2.55 ftl. The values were adjusted among *S*s and between eyes in order to produce suppression phases of approximately equal duration for each eye and of sufficient length for test-probe presentation at all temporal points. Luminance was measured and checked periodically with a Pritchard Spectra photometer (Model 1970-PR). Typically, luminance levels were checked at the viewing port before each experimental session. Monitoring measurements of the rivalry displays within a session were made with a Macbeth illuminometer.

To obtain recognition thresholds under suppression and dominance, any one of four equally probable letter forms (A, N, T, and U) was superimposed on the center disks of the rivalry targets for brief durations. The letters were of pica-style type; each subtended 22′ visual angle at its extreme dimension. These stimuli were constructed of the same materials as the rivalry targets. When transilluminated, the letter appeared as a white outline form against the ground of the rivalry target disk.

The lamp driver and timing units of a modular tachistoscope system (Scientific Prototype Manufacturing Corporation, Model H) were used to transilluminate each letter form for controlled brief durations. Specifically, the lamp driver controlled the duration of a pair of gas discharge lamps (mercury vapor argon with a magnesium tungstate phosphor), the lamp for each eye being mounted in a separate lamp house. These lamps were selected from a large sample on the basis of their firing characteristics (i.e., sharp square wave and reliability). Within each lamp house, a lens collimated and magnified the light beam, the magnification serving to eliminate luminosity gradients in the beam. Directly in front of this lens, a set of neutral density filters was serially aligned. Outside each lamp house and in the optic path of the light beam, the four letter forms were located on a rotatable wheel so that each form could be rotated into position as desired. The positioning of each letter form was accomplished by a detent locking system that ensured that the letter forms would always appear in precisely the same position on every presentation. The duration of each lamp firing was controlled by solid-state timers whose absolute calibration and repeat accuracy were checked and monitored by an electronic counter (Hewlett-Packard, Model 5233L). The wave form for the flashes of both lamps was periodically examined by an oscilloscope (Hewlett–Packard, Model 120A).

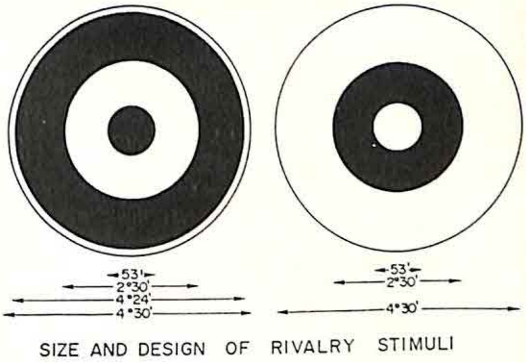


FIG. 2. The achromatic rivalry displays. (Test stimuli appeared in the center of the innermost disks.)

Three solid-state timers were used to provide for three different time delays between the start of a rivalry phase and the presentation of the test form. On a given trial, a single switch (under *E*'s control) selected the appropriate timer and lamp (i.e., eye). Simultaneous with the onset of a lamp, *S* received a brief auditory signal (1,000 Hz., .1 sec.), which informed him that he must make a forced-choice response. An Esterline-Angus event recorder equipped with an auxilliary high-speed drive was used to record the duration of suppression and dominance phases. Two clocks (Hunter klockounters) measured the durations of the one or two rivalry phases involved in every trial.

To report continuously on rivalry alterations and to initiate the presentation of the test stimulus, *S* operated a telegraph-like key whose output was connected both to the event recorder and to the logic circuitry that controlled the tachistoscope lamp system. To report his forces response, *S* used four pushbuttons corresponding to the four letter forms.

A moderate level of noise was delivered to *S* through stereo headphones to mask the sound of rotation of the stimulus wheels and other incidental noises.

*Design and procedure.*—The general design of the experiment was to measure differences in forced-choice form recognition thresholds between suppression and dominance states at three temporal points of onset of the test stimulus within the suppression and dominance phases. The first step was to establish, for each *S* individually, stable recognition thresholds under nonsuppression conditions that would yield correct detections in the range of 70%–80% by manipulation of the duration and luminance of the test stimulus. The range of durations was 30–65 msec.; the range of luminances was 1.4–4.4 ftl. During this phase, the eye under test viewed the rivalry target appropriate to that eye and the test forms were superimposed against the central disk of the rivalry target; the other eye viewed a homogeneous contour-free field whose luminance was the same as the white portions of the rivalry target appropriate for that eye. It should be noted that this binocular combination did not produce patent rivalry; the rivalry target was alway dominant. During the establishment of these thresholds, the conditions simulated in some ways those present during the actual rivalry conditions. Specifically, *S*s initiated their own trials, they remained somewhat uncertain about which eye was going to be stimulated on a given trial, and the a priori probability of occurrence of any one of the four letter forms was .25. In other research, it had been established that these four letters were equally discriminable in terms of the percent of correct recognitions, a finding which was verified during the threshold-setting phase of the present experiment.

The second major task prior to formal data collection was determination of the temporal duration of suppression and dominance phases for each *S*. The objective was the presentation of the test stimulus at (*a*) an early onset point, where the probability of change of state (PCS, probability that a state of suppression would change to a state of dominance, and vice versa) was low, on the order of .05, (*b*) a late point, where the PCS was .50 or greater, and (*c*) a third point that fell somewhere between the early and late points, a so-called middle point.

Estimates of duration were obtained from 60-sec. observation periods during which *S*s reported continuously on the phenomenal status of the 53′ rivalry targets. The rivalry between these disks tends to be unitary, with a negligible amount of incomplete states of dominance and suppression. The *S*s were instructed to assume a passive attitude toward the fluctuations of rivalry and to assume a high criterion for rivalry change, signaling changes only when there was complete dominance or suppression. From the distribution of these durations, decisions were made for each *S* about the most appropriate points for presentation of the test stimuli to represent the early, middle, and late points.

The goal was to present the test stimuli at points within the duration of the suppression and dominance phases that would permit a reasonable test of the relationship between suppression magnitude and duration. Although the duration of any individual rivalry phase is quite unpredictable, the distribution of rivalry phases allows the selection of points of onset that, on the average, represent different probabilities of change of state. Thus, presenting the stimulus immediately after the beginning of a phase assures that the PCS is low (e.g., .05) and provides a measure of the suppression effect during the early portion of a suppression or dominance phase. Sampling the suppression effect during the later parts or phases poses greater difficulty. Although it would have been ideal to select a late point where the PCS was .90, instead of the PCS .50 late point actually used, variability in the duration of suppression and dominance phases would have produced few phases that met that criterion. Further, it seems reasonable to assume that a PCS of .50 would provide an adequate indicator of a temporal trend in suppression magnitude.

The temporal points selected for CM for the early, middle, and late intervals were 560, 1,260, and 2,100 msec. For DC the respective values were 500, 1,100, and 1,700 msec. For KK the values were 1,080, 2,150, and 4,050 msec. These temporal points were based on the intial sample of rivalry durations; the extent to which they met the intended PCS criteria during the experiment was determined by taking additional samples of phase durations throughout the experiment. It became apparent when these additional samples were taken that some departures from the intended PCS values were occurring, but it was decided that these departures were not large enough to warrant changing the points of onset. The actual PCS values used in the experiment, as estimated from the samples of phase durations, were the following: For CM, the early interval values were .01 suppression, .14 dominance; for the middle interval, the values were .21 suppression, .23 dominance; for the late interval, these values were .57 suppression, .57 dominance. For DC the respective values were early, .01 suppression, .01 dominance; middle, .08 suppression, .09 dominance; and late, .30 suppression, .36 dominance. For KK the respective values were early, .10 suppression, .23 dominance; middle, .43 suppression, .42 dominance; and late, .73 suppression, .67 dominance.

Since for the late temporal interval (PCS .50) only half of the presentations could be expected to occur within the appropriate temporal category, it was decided at the outset to accept unequal numbers of stimulus presentations (trials) with the restriction that every temporal interval for suppression and for dominance for each *S* would consist of at least 50 trials. To meet this requirement, the frequency of presentations of stimuli for the three temporal intervals was adjusted throughout the experiment. Further, the procedures were designed to save presentations that fell in an inappropriate temporal interval due to a spontaneous change of rivalry state; for example, a presentation intended for late suppression might actually occur in the middle interval of the subsequent dominance phase and therefore was counted in the latter category. Such presentations could be classified because on each trial *S*s reported on two successive phases of rivalry whose durations were sequentially recorded on the clocks.

The general procedure on each trial was as follows. The *S* continuously reported, by means of his response key, on the phases of rivalry. Following a predetermined schedule, *E* selected the rivalry state and the temporal interval in which the stimulus would be presented. Upon presentation of the stimulus, *S* stopped reporting on rivalry and immediately gave a forced-choice response about the letter that had been presented. In addition, *S* indicated the actual state of rivalry present at the moment the stimulus was presented. Trials were accepted only if there was agreement between *S*'s poststimulus report and the state of rivalry signaled by his response key. The number of trials discarded because of disagreement between these two methods of report proved to be quite small. During a single session, each of the four letter forms was randomly scheduled to appear at least three times under each of the three temporal intervals for both the suppression and dominance phases (2 × 3 × 3 × 4). On a given trial, *S* was undertain about which rivalry phase was to be stimulated and which temporal interval was to be employed. To introduce uncertainty as to which eye was going to be stimulated, an additional 18 trials were presented to the eye not under test, these being delivered in random sequence. The intent of keeping *S* uncertain about the condition of a specific trial was to minimize potential attitudinal factors that may have favored one variable over another.

Although a prototypical or standard session consisted of 90 trials, the actual number of trials per session varied somewhat in order to provide a sufficient number of trials for all the temporal intervals and in order to maintain an equivalent number of letter-form presentations. The total number of trials (excluding trials presented to the eye not under test) obtained over six sessions and the range of trials representing a particular Rivalry × Temporal Interval are as follows: For CM, there were 227 suppression trials and 204 dominance trials, ranging 52–107; for DC there were 226 suppression trials and 231 dominance trials, ranging 73–83; for KK there were 209 suppression trials and 221 dominance trials, ranging 55–124. The intertrial interval averaged 10 sec.; after 45 trials *S*s were forced to rest for at least 10 min.

## RESULTS

In Fig. 3, the percent correct recognition for suppression and dominance for each of the temporal intervals is plotted for each *S* separately. The abscissa is the PCS for the three intervals for each *S*.

The data of each *S* were analyzed separately by an analysis of variance, the unit of analysis being the proportion correct within each of the six sessions for each of the temporal points and conditions of rivalry state. A test of homogeneity of variance ($F_{\max}$ test) revealed that the data for none of the *S*s was significantly heterogeneous. The terms tested by the analysis were sessions, rivalry state (suppression, dominance), temporal interval, and the Temporal Interval × Rivalry State interaction. For each *S* the only significant term was the rivalry condition: For CM, $F\ (1,\ 5) = 19.10$, $p < .01$; for DC, $F\ (1,\ 5) = 17.09$, $p < .01$; for KK, $F\ (1,\ 5) = 53.83$, $p < .001$. For none of the *S*s did the temporal interval or the Temporal Interval × Rivalry Condition interaction terms approach significance (i.e., $F \leq 1$).

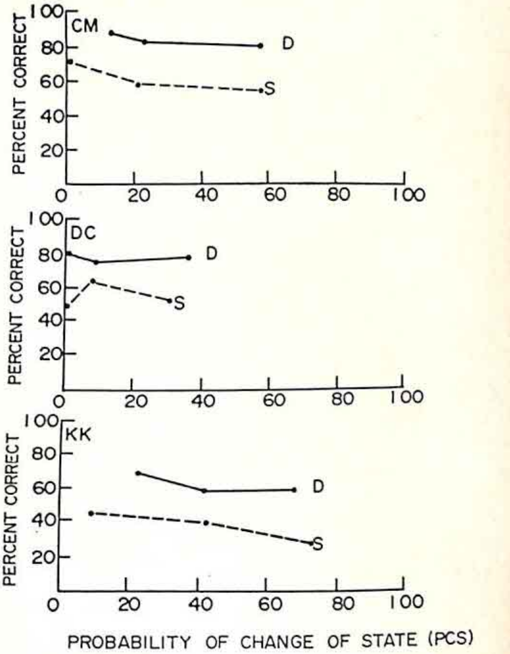


FIG. 3. Mean percent correct recognition responses for each of the three temporal intervals for each *S* under conditions of dominance (D) and suppression (S).

An inspection of the graphs in Fig. 3 reveals that for two *S*s there is some tendency for the suppression effect to increase for the late temporal interval. For CM, the difference in recognition between suppression and dominance increases by 10% from the early to the late temporal interval; for KK this increase was 8%. To determine if this trend approached significance, an additional analysis of variance, using only the early and later temporal intervals, was performed on the data of each *S*. This analysis revealed that while the suppression effect remains significant, there is no suggestion of significance for the temporal interval variable nor for the Temporal Interval × Rivalry Condition interaction. The *F* values for the rivalry condition variable are as follows: for CM, $F$ $(1, 5) = 7.49$, $p < .05$; for DC, $F$ $(1, 5) = 18.43$, $p < .01$; for KK, $F$ $(1, 5) = 30.07$, $p < .005$. For the temporal interval term: For CM, $F < 1$; for DC, $F < 1$; for KK, $F$ $(1, 5) = 2.18$, $p < .20$. For the interaction term, $F < 1$ for all *S*s. Both the analyses using all three temporal intervals and the analyses using two temporal intervals were performed on the untransformed portions obtained within each session. Identical analyses were performed on the data after an arc-sine transformation; the results were essentially identical to those just described.

To determine if differences in discriminability of the letter forms and differences in the proportion of letter forms presented could have played any role in the results obtained from each *S*, a complete analysis of the occurrence of each letter form under each of the conditions of temporal interval and of rivalry condition was made for each each *S*. These analyses revealed that the letter forms were equally discriminable among themselves for both dominance and suppression and, further, that each *S* received approximately the same number of letter forms over all conditions.

With respect to the decline of performance at the late temporal interval, which was exhibited by CM and KK, even though it is not significant it is possible that the decline can be attributed to a lowering of attention. All *S*s reported that they felt most alert at the onset of a phase and then less so as the phase continued unbroken by the presentation of a test stimulus.

## DISCUSSION

There is no doubt that recognition performance under suppression is significantly impaired relative to performance under dominance. The magnitude of the impairment is similar to that observed earlier for the same kind of task by Fox and Check (1966). It is of interest that the magnitude is similar to that found by Eriksen and his colleagues (e.g., Eriksen & Collins, 1965) for visual backward masking using a forced-choice letter-recognition paradigm very similar to the one employed in this experiment.

With respect to the theoretical implications of these data, the version of the correlative hypothesis represented by Wave Form *a* and the transitory hypothesis represented by Wave Form *d* receive no support. Both hypotheses require a decline in the magnitude of the suppression effect, and this is definitely not the case, even when the phase durations are of sufficient length that the probability of termination of suppression is greater than .50, the case of CM, or of .75, the case of KK.

With respect to the correlative hypothesis represented by Wave Form *b* and the constant hypothesis represented by Wave Form *c*, the decision between them is not quite so unequivocal. For, in accord with the assumption of Wave Form *b*, there is some suggestion of an increase in suppression with an increase in duration for two *S*s. But for both of these *S*s the increase is not statistically reliable nor is the increase absolutely large. It is likely that the present experiment could have reliably detected differences in recognition performance of 20% or more and, although a quantitative formulation of the correlative hypothesis represented by Wave Form *b* is not available, it seems plausible to suppose that any reasonable version of that model would have to predict increases greatly in excess of 20%, at least for the latest temporal interval sampled.

These considerations rend the correlative hypothesis of Wave Form *b* so tenuous that it must be rejected. The data from all three *S*s indicate that the suppression effect remains constant throughout the course of the suppression phase, an outcome completely consistent with the key assumption of the constant hypothesis of Wave Form *c*. But a note of caution is in order, in that acceptance of the constant hypothesis is tantamount to accepting the null hypothesis, and the problem of accepting the null hypothesis by virtue of an insensitive experiment has been thoughtfully discussed (Binder, 1963; Grant, 1962). Further, the present experiment has employed only three temporal intervals, and more than three would certainly have been desirable. Given these qualifications, the data fit the assumptions of the constant hypothesis so well that it is necessary to conclude that the constant hypothesis must be accepted.

The rejection of the correlative hypotheses means that the magnitude of suppression and the duration of suppression cannot be accounted for by a single mechanism. It is particularly unfortunate that no evidence was found to support the correlative hypothesis of Wave Form *a*, which is the one that would be anticipated on the basis of a reciprocal inhibition model of rivalry. Indeed, a priori, the decay of inhibition over time would seem to be the most reasonable way for an inhibitory process like rivalry to behave. But the present data make it clear that this is not the way the rivalry mechanism operates, and consideration must now be given to models of rivalry that employ separate mechanisms for the control of rivalry phase durations and for the magnitude of the suppression effect.

## REFERENCES

BINDER, A. Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 1963, **70**, 107–115.

ERIKSEN, C. W., & COLLINS, J. F. Reinterpretation of one form of backward and forward masking in visual perception. *Journal of Experimental Psychology*, 1965, **70**, 343–351.

FOX, R., & CHECK, R. Forced-choice form recognition during binocular rivalry. *Psychonomic Science*, 1966, **6**, 471–472.

FOX, R., & CHECK, R. Detection of motion during binocular rivalry suppression. *Journal of Experimental Psychology*, 1968, **78**, 388–395.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psycgological Review*, 1962, **69**, 54–61.

LEVELT, W. J. M. *On binocular rivalry.* Soesterberg, the Netherlands: Institute for Perception RVO-TNO, 1965.

WALES, R., & FOX, R. Increment detection thresholds during binocular rivalry suppression. *Perception & Psychophysics*, 1970, **8**, 90–94.

(Received August 16, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 290–296

# MULTIDIMENSIONAL STIMULUS CONTROL: EFFECTS OF TRAINING AND/OR TESTING[1]

FREDERICK L. NEWMAN,[2] C. FRANK ANDREONE, LYNNE WASHBURN, AND RONALD B. PURTLE

*University of Miami*

In a nine-group experiment, 126 human *S*s were employed to assay overshadowing and masking effects as a function of differential training and testing conditions. The training conditions contrasted two counterbalanced groups differentially trained along one of the two dimensions (the other dimension present, but nondifferentially reinforced), with a third group receiving differential training along both dimensions simultaneously. Testing conditions contrasted single-dimension testing along one of the two dimensions (in the presence of the S+ component of the other dimension) with simultaneous testing along both dimensions. Simultaneous testing resulted in flatter gradients overall. Joint effects of training and testing were observed where some effects were shown to develop either early or late over the course of five alternating training and testing phases.

A major purpose of multidimensional stimulus control studies has been to investigate how the components of the discriminative stimulus exert control in combination. One group of studies, summarized by Cross (1965) and by Purtle and Newman (1969), have attempted to test some models describing the combined stimulus control of two dimensions. The Purtle and Newman study showed that the best descriptive rule of combined stimulus control was dependent upon the nature of the training procedure and the relative saliency of the two stimulus dimensions. Unless a stimulus feature of low saliency was specifically differentially reinforced during training, the more salient feature would overshadow the development of control by the less salient feature. The evaluation of these effects was derived from a testing situation during which each of the two dimensions were simultaneously varied. The simultaneous testing procedure was necessary for testing the models of combined stimulus control.

Several investigators have warned that the design of a test of stimulus control could contaminate a view of the training effects. In pigeon operant studies, the presence of a salient component of the S+ complex during testing along the dimension of a less salient feature masks the existence of control by the less salient component (Thomas, Svinicki, & Svinicki, 1970). Earlier articles had suggested that the structure of the testing situation may also modify control via learning derived from the experience of testing itself (Prokasy & Hall, 1963). Most of the studies investigating testing effects have used infrahuman *S*s. Furthermore, most infrahuman investigations have employed single-dimension tests in the presence of another salient feature of S+, rather than the simultaneous testing procedure.

The present experiment investigates the effects of using a single-dimension test as contrasted with a simultaneous test procedure in evaluating combined stimulus control Single-dimension testing along one dimension took place in the presence of the S+ component of the second dimension. Three training conditions preceeded each test: differential training where two dimensions were redundantly relevant; differential training along the more salient dimension with the less salient component constant; and differential training along the less salient dimension with the more salient component constant over positive and negative instances. Because of the imbalance in the saliency of the two dimensions, the design allowed for the evaluation of the "overshadowing" effects of training, the "masking" effects of testing, as well as a possible interaction.

---

[1] This research was reported at the May 1971 meeting of Rocky Mountain Psychological Association at Denver and prepared while the first author was on leave at New Mexico State University and The Center for Advanced Study in the Behavioral Sciences.

[2] Requests for reprints should be sent to Fred Newman, Box 8185, Department of Psychology, University of Miami, Coral Gables, Florida 33124.

Another feature of the present study was that the negative training stimulus was fixed for all *S*s rather than varied over non-S+ points along the two dimensions. In the Purtle and Newman (1969) study, a randomly varied S− was used between *S*s. By randomly varying S−, Purtle and Newman avoided inferential problems related to possible behavioral contrast or peak shift effects. By fixing the location of S− for all *S*s in the present study, the intention was to evaluate these effects with humans in the instrumental training and testing situation.

## METHOD

*Subjects.*—The *S*s were 32 female and 94 male undergraduate students in the introductory psychology course at the University of Miami. While participation was voluntary, it was rewarded with the possibility of improvement in course grade.

*Designation of stimuli.*—As was the case in the earlier study (Purtle & Newman, 1969), the two stimulus dimensions were angular orientation of a rectangular aperature and tactile texture of sandpaper contained within the area of the rectangle. Specifically, when sandpaper discs of differing grit levels were held up behind at 12.7 × 101.6 mm. (½ × 4 in.) aperture varying in angular orientation, a blindfolded *S* would feel a textured rectangle which could vary in either angular orientation, or texture, or both. The results of the Purtle and Newman study, and the verbal reports of *S*s in the present study, showed that the tactile texture cue was more salient than the angular orientation cue.

The five grit levels of sandpaper were selected because of their accessibility at local hardware stores (3M brand, Grit No. 50, 80, 120, 220, and 320). Thirty *S*s, not used in the regular experiment, were given the task of psychophysically equating distance along the dimension of angular orientation to perceived distance of Grit No. 50, 80, 220, and 320 from the Grit No. 120 sandpaper. The standard position of angular orientation was 45° clockwise, where adjustments toward the horizontal represented "rougher." The method of adjustment was employed and all adjustments were restricted to a single quadrant. The psychophysical task was a replication of the one done at another university by Purtle and Newman (1969). The results for the replication were (corresponding to the five grit numbers, low to high): Angle 1, 23.3°; Angle 2, 30.6°; Angle 3 (the standard), 45°; Angle 4, 64°; and Angle 5, 71.4°. While there was some overlap among the four 95% confidence intervals, 4 of the 30 *S*s accounted for the variability that produced the overlapping intervals.[3] Thus, there were five textures ($T_1$ to $T_5$ representing grit numbers low to high) and five angles ($A_1$ to $A_5$) for which simultaneous tests of stimulus control provided 24 stimulus pairs representing the two-dimensional array.

*Apparatus.*—The sandpaper discs were mounted on 6.35 × 127.3 × 127.3 mm. (¼ × 5 × 5 in.) plywood panels. The discs were held up to and covered the 12.7 × 101.6 mm. rectangle cut out of the center of a standard 2-lb. plastic coffee can lid mounted onto a 2-lb. coffee can with both ends removed (see Purtle & Newman, 1969). A white-noise unit was used to mask any discriminating sounds derived from movement of the sandpaper blocks.

*Design and procedure.*—Forty-two *S*s were assigned to one of three training conditions: specific discrimination training to angular orientation (texture $T_3$ present but nondifferentially reinforced); specific discrimination training to texture (angularity, $A_3$, nondifferentially reinforced); and nonspecific discrimination training where both texture and angularity were differentially reinforced. The S+ compound was $T_3A_3$ for all groups. The S− compound differed among groups, but not within the groups. When training was specific along the angularity dimension, the S− compound was $T_3A_2$ or $T_3A_4$ (counterbalanced between half of the 42 *S*s). Note that the texture value remains constant at $T_3$ over S+ and S− instances for angle specific training. Likewise, for specific discrimination training to texture, the S+ compound was $T_3A_3$ and S− was either $T_2A_3$ or $T_4A_3$ (counterbalanced between *S*s). The nonspecific discrimination training *S*s had $T_3A_3$ as S+ and either $T_2A_2$ or $T_4A_4$ as the S− compound (again counterbalanced between *S*s). The possible compounds of $T_2A_4$ and $T_4A_2$ were not used as S− instances. Table 1 summarizes the design.

During a training phase, the blindfolded *S* was told to say YES whenever he "noticed this" test stimulus (directing *S* to reach his fingertips of his preferred hand forward in the cylinder and then to return his fingertips to the near edge) and to say NO whenever any other test stimulus was presented. An example of a NO stimulus was then given. This was repeated three more times. "Remember, this is the stimulus you say 'Yes' to, . . . and this is an example of a 'No' stimulus." No mention of either dimension was made in the instructions.

A test phase immediately followed the training phase. There were three testing conditions all con-

---

[3] There was a disparity between the results of the psychophysical scaling study at Miami and that found in New Mexico by Purtle and Newman (1969). Specifically the two angles called "rougher" were 8.1° and 25.1° in the earlier study. We have no explanation for the discrepancy of results here. The stimulus judgments of the angles representing the "smoother" direction were consistent with the earlier results of 62.7° and 72.1°, respectively.

TABLE 1

EXPERIMENTAL DESIGN SPECIFYING THE NINE COMBINATIONS OF THREE TRAINING AND THREE TEST CONDITIONS

| Training cond. | Test cond. | | |
|---|---|---|---|
| | Angle varies $T_3$ constant | Texture varies $A_3$ constant | Both vary |
| Angel discrimination<br>$T_3A_3$ vs. ($T_3A_2$ or $T_3A_4$)[a] | Training (A)<br>Test (A) | Training (A)<br>Test (T) | Training (A)<br>Test (TA) |
| Texture discrimination<br>$T_3A_3$ vs. ($T_2A_3$ or $T_4A_3$)[a] | Training (T)<br>Test (A) | Training (T)<br>Test (T) | Training (T)<br>Test (TA) |
| Nonspecific discrimination<br>$T_3A_3$ vs. ($T_2A_2$ or $T_4A_4$)[a] | Training (TA)<br>Test (A) | Training (TA)<br>Test (T) | Training (TA)<br>Test (TA) |

[a] Counterbalanced between *S*s.
Note.—A = angle, T = texture.

sisting of the sequential presentation of 25 test stimuli. For the simultaneous test condition, the 25 combination pairs of the five textures and five angles were each presented once in random sequence. For the single-dimension tests on angular orientation, each of the five angles was paired with $T_3$ five times, given in a random sequence of 25 test stimuli. Likewise, for single-dimension tests along the texture dimension, each of the five grits of sandpaper was paired with $A_3$ five times in a random sequence of 25 test stimuli.

After *S* went through the test phase, a 2-min. rest period followed. An *S* was then told that *E*s wanted to refresh his memory about the nature of the YES and NO stimuli, and then the training phase followed by the testing phase was repeated. Overall, each *S* experienced five relpications of training and testing.

## RESULTS

The measure of *S*'s tendency to respond to a particular test stimulus compound was defined as the number of YES responses out of the five times which that stimulus component was presented in a random sequence of 25 presentations of test stimulus compounds. The scale for a single trial block of 25 stimulus presentations was, therefore, 0 to 5. The average number of YES responses on the first two blocks of 25 test trials were analyzed to represent the initial effects of training and testing. The response average on the last three blocks of 25 test trials were analyzed to represent asymptotic performance. The synthesis of the data proceeds by first inspecting the asymptotic results, Trial Blocks 3–5, and then to inspect transitory effects shown in the analysis of the first two trial blocks.

*Last three trial blocks.*—There were no significant differences in response averages among any of the group partitions. All differences were reflected by analyses performed on between-group quadratic trends.[4] Because of this, the same information can be gleaned from the absolute gradients as can be found in the relative gradients. The absolute gradients of group averages over the last three test trial blocks are presented in Fig. 1.

The overall effect of testing simultaneously along both dimensions was to produce a flatter graident than under single-dimension testing in the presence of the S+ component from the untested dimension. The interaction of the quadratic trend components under the two test conditions was significant for the angular orientation dimension, $F\ (1, 72) = 17.809$, $p < .001$, and for the texture dimension, $F\ (1, 72) = 23.162$, $p < .001$.

Under the simultaneous testing condition, the analysis of the last three trial blocks showed no differential training effects. What was explicit in the results of the simultaneous tests was the relative saliency of the two stimulus dimensions under test. The overall quadratic trend over the five texture stimuli under the simultaneous test condition was significant, $F\ (1,\ 72) = 41.05$, $p < .001$, and accounted for approximately 25% of the total within-*S* variance under the simultaneous test condition. While there was a significant quadratic trend along the angularity dimension under simultaneous testing, $F\ (1,\ 72) = 3.95$, $p < .05$, the trend accounted for less than 1% of the total, simultaneous test, within-*S* variance. Differences between training group quadratic trends under simultaneous testing along either dimension were all nonsignificant and collectively accounted for less than .3% of the total within-*S*s variance.

---

[4] All statistical analyses of variance and trends were performed within one stimulus dimension and not across the two dimensions. While the precaution of psychophysical analysis may be a reasonable tact, we were simple not convinced that a method of adjustment would result in psychological distance functions which are isomorphic with the results of our generalization tests. There is certainly a long history of data suggesting that a units (hypothesis) effect would contaminate such distance functions in test of stimulus generalization (Mednick & Freedman, 1960).

Single-dimension testing results, in the presence of the S+ component from the other dimension, appear in the upper two panels of Fig. 1. The differential extent of control along each of the two dimensions was not as marked in the results of the single-dimension testing as it was in the simultaneous test. The overall quadratic trend across texture test stimuli was significant, $F\ (1, 72) = 172.15$, $p < .001$, and accounted for about 28% of the total (single-dimension test) within-*S* variance. Across the angular orientation dimension, the quadratic trend was significant under the single-dimension test, $F\ (1, 72) = 63.66$, $p < .001$, and accounted for about 15% of the total within-*S* variance.

One may be tempted to accentuate the differences in percent of total (single dimension) within-*S* variance between the texture and angle quadratic trends. The difference developed because another fixed effect was sharing the total within-*S* (angle test) variance along with the overall quadratic trend. Specifically, the single-dimension angularity test of quadratic trends showed a significant difference among the three training group trends, $F\ (2,\ 72) = 24.44$, $p < .001$. The *S*s trained specifically to texture and tested along the angle dimension showed the flattest, although significant, gradient among the hree groups. The contrast of the texture

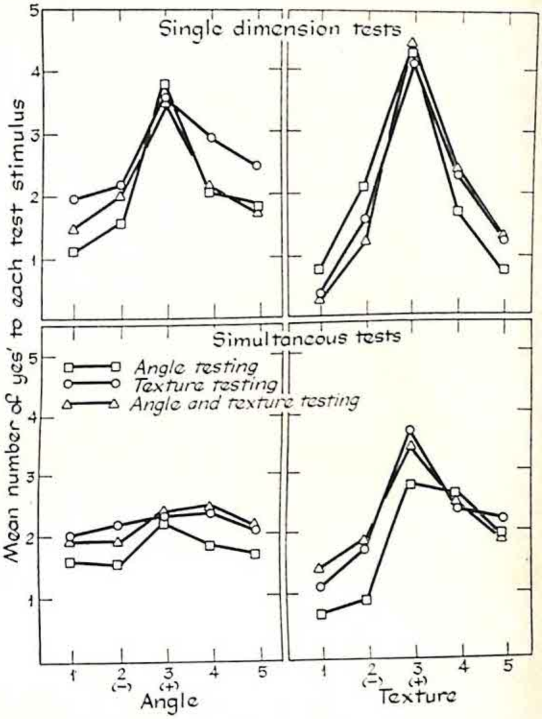


FIG. 1. Mean number of YES responses to each test stimulus over last three test trial blocks, presented in a 2 × 2 display; Single-Dimension Testing (upper two panels) and Simultaneous Testing (lower two panels) × Angle (left panels) and Texture (right panels).

trained group with the other two groups was significant, $F\ (1,\ 72) = 13.15$, $p < .001$. The proportion of total within-*S*s variance accounted for by the Training Groups × Quadratic Test Stimuli interaction was about 13%. These fixed effects together accounted for approximately 28% of total within-*S* variance on the angularity dimension, which was the same estimate of proportion found for the quadratic trend along the texture dimension.

*Trial Blocks 1–2.* The test results along the testure dimension on the first two trial blocks paralleled that which was observed on the last three trial blocks. The simultaneous testing procedure produced flatter gradients than did the single-dimension test procedure, $F = 7.037$, $p < .01$. When contrasting the overall texture quadratic trends observed on the first two trial blocks with those observed on the last three trial blocks, a significant increase in quadratic components was found, $F$ (1, 72) = 4.759, $p < .05$. Finally, with regard to the more salient texture results, no training group effects or Training Group × Test Procedure interactions were observed during the first two trial blocks, or across trial blocks (see Fig. 2).

For the less salient feature, angular orientation, several effects varied as a function of trial blocks. While there was a general steepening of the texture gradients over trial blocks, no such overall tendency was observed along the angle dimension, $F$ (1, 72) = 1.483. There were, however, some subgroup changes over trial blocks. The most important change in quadratic trends was observed under the single-dimension angle test. The quadratic trend component of *S*s specifically trained to texture remained about the same over trial blocks, while there was an increase in the magnitude of the quadratic trend components for the two angle training groups. The specific contrast of the quadratic trend components reflecting these relationships was significant, $F$ (1, 72) = 5.103, $p < .05$. Thus, the training groups effect observed on the last three trial blocks under the conditions of the single-dimension test were not observable during the initial trial blocks.

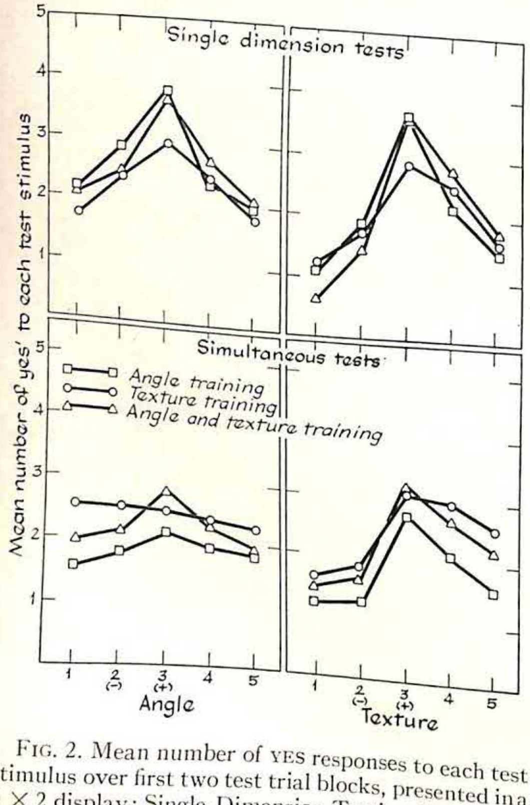


FIG. 2. Mean number of YES responses to each test stimulus over first two test trial blocks, presented in a 2 × 2 display: Single-Dimension Testing (upper two panels) and Simultaneous Testing (lower two panels) × Angle (left panels) and Texture (right panels).

Under the conditions of the simultaneous test, no increases in the angle gradients were observed. In fact, the slight but reliable angle gradient observed by the redundant-cue (angle and texture) trained *S*s in the first two trial blocks, $F$ (1, 72) = 4.779, $p < .05$, flattened and was not significant for the last three trial blocks, $F$ (1, 72) = 1.368.

## DISCUSSION

A prospective of how the present results compare with the data of the Purtle and Newman (1969) study ought to preceed a discussion of theoretical implications. The gradients observed under the simultaneous testing condition in the present study appear to be much flatter than those observed in Purtle and Newman. The flatter gradients were very likely due to the fixing of S− one unit from S+ for all *S*s in all groups, whereas S− in the earlier study could have been one or two units removed from S+. In any event, the combined effect of the present study's training procedure, common to all groups, and the simultaneous testing procedure results in flatter gradients with no reliable effects of training. The reliable training effects were only observed under single-dimension testing of control by the less salient angle feature. Overall, the results showed that there can be a fragile balance among the factors of cue saliency and testing procedures. The fragility of such balance recommends that multidimension stimulus control model testing may be even more difficult than Sutherland had suggested (Sutherland & Mackintoch, 1971).

While the present study was not designed to investigate the models discussed by Purtle and Newman (1969), the test of the models on the simultaneous test results was informative. The dominance model provided the best description of combined stimulus control for all three training groups. The earlier study found that specific angle training overcame the imbalance of cue saliency and allowed the angle feature to compete with texture. The training procedure used in the present investigation apparently precluded the development of angle contol in the simultaneous testing group. There may have also been some effect of differences resulting from a change in the psychophysical angle–texture scale (see Footnote 3). Post hoc analyses did not reveal any Counterbalancing × Training effects.

The difficulty in establishing stimulus control in the present study, relative to Purtle and Newman (1969), provides some interesting insights regarding the interplay of cue saliency and training procedures which were not afforded in earlier efforts. The exposition of cue saliency and training procedure interaction was best explored by using the results of the more salient dimension, texture, as the referent of performance for the less salient angular orientation dimension.

The texture cue was certainly the more salient feature of S+ and, as one might expect, texture control was not differentially affected by any one of the specific training procedures. While the busier simultaneous testing procedure homogeneously flattened the texture gradient, the pervasive control by the S+ texture component was never in doubt. The progressive steepening of the texture gradients from early to late trial blocks, under both testing procedures, suggested a general effect demanded by those aspects of the training and testing procedures common to all groups.

Upon inspecting the result of tests along the less salient angle dimension, some groups showed performances similar in form to the testure results and others did not. First, significant angle training effects were only observed under the single-dimension angle tests. As such, one can conclude that the simultaneous testing condition did "mask" training effects. From a methodological viewpoint, decisively strong training differences would be needed to overcome possible masking effects of a simultaneous test procedure.

The most interesting results were the differential effects of the training procedures on control by the less salient angle feature under the conditions of the single-dimension angle test, with the S+ texture component present. Two groups showed a steeping of the angle gradients similar to that which was observed for the more salient texture gradients. The two groups showing the increased level of angle control from early to late trial blocks were those *S*s trained specifically to angle and those *S*s trained redundantly along both angle and texture dimensions. The parallel of angle and texture gradients in these two cases was not unexpected. Specific training to angle along with the repeated angle testing experience was probably sufficient for the angle cue to exert increased control paralleling the increased control exhibited by the more salient texture cue.

Angle control by *S*s specifically trained to texture (angle S+ feature constant during training) was exhibited under the single-dimension angle tests (texture S+ feature constant during testing). For the initial test trial blocks, the angle gradients of the texture trained *S*s was not noticeably different from the angle trained *S*s. The differential training effect for the single-dimension angle testing data was statistically significant over the latter part of testing. What was observed was a steepening of the angle gradients by the specific angle training groups, and not by the specific texture trained *S*s. The relative stationary maintenance of angle control by texture trained *S*s was most likely a function of the angle testing procedure itself rather than by the nondifferential angle training procedure. It was likely that a number of additional angle test trial blocks would have steepened the angle gradients for the specific texture trained *S*s; but that would only be because of refinement of that which was initially learned during single-dimension angle testing. For a more salient cue (e.g., texture), differential training or testing would not be as crucial. Indeed, the results of the present experiment suggested that such factors were not crucial for the establishment of control by the most salient texture feature.

For the less salient feature, angle in the present case, the development of control can occur more effectively under a differential training procedure. A less efficient vehicle of establishing stimulus control, but nevertheless effective, would be by employing single-dimension testing of control by the less salient cue in tandum with nondifferential reinforcement. In terms of the present human *S* experiment, the alternation of the nondifferential angle training (and differential texture training) procedure with single-dimension angle testing very likely alerted *S*s to the relevance of the angle feature of S+. As such, a reasonable explanatoin of the present results could be developed from the hypothesis learning construct proposed by Levine (1971). On the other hand, an animal selective attention model, such as the one described by Sutherland and Mackintosh (1971) can also be employed as a theoretical base of explanation. In order for the Sutherland and Mackintoch model to be effective in the present case, however, the model must integrate the structure of the tests of stimulus control as an integral art of the learning and/or stimulus selection process. Within the context of the present experiment, joint consideration of both the training and the testing procedures as part of learning processes was clearly reasonable. We would also maintain, as Prokasy and Hall did in 1963, that the same integration is required in most animal studies of multidimensional stimulus control. The segregation of overshadowing as a training effect and masking as a testing effect, in this light, is probably inappropriate.

## REFERENCES

Cross, D. V. Metric properties of multiple stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization.* Stanford, Calif: Stanford University Press, 1965.

Levine, M. Hypothesis theory and nonlearning despite ideal S–R Reinforcement contingencies. *Psychological Review*, 1971, **78**, 130–140.

Mednick, S. A., & Freedman, J. L. Stimulus generalization. *Psychological Bulletin*, 1960, **57**, 169–200.

Prokasy, W. F., & Hall, J. F. Primary stimulus generalization. *Psychological Review*, 1963, **70**, 310–322.

Purtle, R. B., & Newman, F. L. Multidimensional stimulus generalization of a tactile response along the dimensions of angularity and texture. *Journal of Experimental Psychology*, 1969, **81**, 566–571.

Sutherland, N. S., & Mackintoch, N. J. *Mechanisms of animal discrimination learning.* New York: Academic Press, 1971.

Thomas, D. R., Svinicki, M. D., & Svinicki, J. G. Masking of stimulus control during generalization testing. *Journal of Experimental Psychology*, 1970, **84**, 479–482.

(Received August 17, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 297–301

# EFFECT OF A FORWARD INDICATOR ON BACKWARD MASKING

TERRY J. SPENCER,[1] LARRY HAWKES,[2] AND GREGORY MATTSON

*Kent State University*

Backward masking functions were obtained for a single-letter (SL) visual display, and for 12-letter (12-L) circular displays presented with either a simultaneous or 150-msec. leading indicator. The *S*s were required to report both the indicated letter (first position) and the letter directly opposite (second position) in the 12-L conditions. The first-position forward indicator function asymptoted to its No-Mask Control approximately 100 msec. faster than the first-position simultaneous indicator functions, and at about the same mask delay as the SL function. The pattern of results suggest that differences in masking ranges between single- and multiple-element displays are a function of encoding strategy rather than a difference in icon formation time.

It is a common finding that backward masking functions with a pattern mask have a greater range with multiple-element as compared to single-element visual displays (Averback & Coriell, 1961; Spencer, 1969; Spencer & Shuntich, 1970). This difference in masking range is usually assumed to be a function of the encoding strategy appropriate for each type of display. For example, the location and identification of the target letter in a single-letter display is presumed to occur much faster than in a multiple-letter display with indicator, and the masking range is assumed to represent the processing time in each case. Evidence provided by Eriksen and Collins (1969), and recently confirmed by Eriksen and Hoffman (1972), suggests that approximately 100 to 200 msec. are required to encode the meaning of an indicator and direct attention to a critical item. An indicator presented 200 msec. forward of a multiple-letter display improved performance as compared to a simultaneous indicator. In addition, Spencer has provided evidence that encoding from iconic storage is dependent on an indicator in a multiple-letter display. Delaying an indicator 100 msec. increased the masking range for a multiple-letter display by a comparable amount.

However, it has also been suggested that the difference in masking range between single- and multiple-element displays may be due, at least in part, to a difference in icon formation time.[3] The fact that multiple-letter displays typically have a number of elements in close proximity may retard the rate of contour formation relative to that of a single-element display. The greater masking range with the multiple-element display may thus reflect the additional time required for development of the icon prior to its encoding.

The present experiment obtains backward masking functions for 12-letter (12-L) circular displays with both simultaneous and forward indicators, as well as for a single-letter (SL) display. If the difference in range between SL and 12-L masking functions is primarily a result of differences in icon formation time, then little or no reduction in the masking range for a 12-L display should be expected with a forward indicator. However, if a 12-L display's icon is developed at or about the same rate as that of an SL display, and a forward indicator can serve to direct focal attention prior to the 12-L display's onset, the backward masking range with a forward indicator may approximate that of an SL, or at least be substantially shorter than a 12-L display with simultaneous indicator.

## METHOD

*Subjects.*—Five *S*s (three males) with normal or corrected-to-normal vision were recruited from the

---

[1] Requests for reprints should be sent to Terry J. Spencer, Department of Psychology, Kent State University, Kent, Ohio 44242.

[2] Now at the University of Utah.

[3] I am thankful to Professor Harold W. Hake for pointing out this possible interpretation.

Kent State University population and paid for their services.

*Apparatus and stimuli.*—Stimuli were presented in a three-field Scientific Prototype Model GB tachistoscope. The test stimulus was presented in Field 1, the indicator in Field 2, and the pattern mask in Field 3, all fields being front lighted. Present in the center of Field 3 was a small dim back-lighted fixation cross subtending a visual angle of .2°. The stimuli consisted of the letters A, O, V, T (black Paratipe No. 11316). For the 12-L conditions, each of the 4 letters were represented three times on each of 12 white stimulus cards. The 12 letters were arranged in a circular pattern corresponding to the positions of a clock and were viewed at a distance of 127 cm. The diameter of the circle subtended a visual angle of 2.0°, and the letters subtended a visual angle of .2°. The letters were assigned at random to each card with the restrictions that the same letter was never adjacent to itself and that each letter occupied all 12 clock positions equally often over the 12 cards. There were 48 SL stimulus cards representing all possible combinations of clock positions and letter types. This latter set was randomly split into two smaller sets containing 24 cards with the restriction that each letter and each position were equally represented in each set. There were 12 indicator cards, 1 for each clock position of the test stimulus. Each indicator was bar shaped (.12° × .88°), white against a flat-black background, with the close edge located .47° from the outside edge of the test form letters. The pattern mask was constructed by placing four Paratipe Ws back to back and side by side, partially overlapping, to form a pattern similar to three adjacent Xs. This pattern was duplicated for each of the clock positions of the test stimulus, all on 1 card. A flat-black card with a circular hole (diameter subtending 2.8°) was placed in front of the test stimulus card and the pattern mask. This allowed the circular test field and the pattern mask to be viewed, while providing a dark background for the outside indicator in Field 2, which was clearly visible in all exposure conditions. A flat-black circular card with diameter of 2.8° (the "hole" from the card placed in front of Field 1) was suspended by black threads over the center of Field 2. This eliminated luminance summation (Eriksen, 1966) from the center of the indicator card in Field 2 with the stimulus in Field 1. A 2.0 neutral density filter was placed in front of Field 1 and a .8 filter in front of Field 2. The luminances, as measured by a MacBeth illuminometer with a white card in each field, were .38, 3.5, and 23 mL. for Fields 1, 2, and 3 respectively, with the above filters in place.

*Procedure.*—Backward masking functions were obtained for one condition of SL display with stimulus onset asynchronies (SOAs) of 50, 100, and 150 msec., and three conditions of 12-L display, each with SOAs of 100, 150, 200, 250, and 300 msec. A No-Mask Control was also obtained for each condition. The four experimental conditions were as follows:

1. A single letter (SL) presented without an indicator with display energy set for about a 70% No-Mask Control recognition accuracy (uncorrected for chance guessing).

2. Low Energy-Simultaneous Indicator (LE-SI)—12-L display, with energy set for about a 70% No-Mask Control recognition accuracy, presented simultaneously with an indicator.

3. Low Energy-Forward Indicator (LE-FI)—an indicator presented 150 msec. prior to the onset of a 12-L display with the same energy as in the LE-SI condition. For only this condition, the onset of the 12-L display was delayed 150 msec. after *S* initiated a trial.

4. High Energy-Simultaneous Indicator (HE-SI)—an indicator presented simultaneously with a 12-L display whose energy was increased to provide a No-Mask Control performance level equal to or above that obtained in the LE-FI condition. This condition was to control for the possibility that simply increasing the Mo-Mask Control performance level is sufficient for reducing the masking range. If an increase in the No-Mask Control performance in the LE-FI condition is actually due to faster encoding, as opposed to some other factor concommitant with a forward indicator, the masking range in LE-FI should also be shorter than that obtained in HE-SI.

Energy was defined as stimulus duration and averaged 10, 16, 16, and 20 msec. for Cond. 1–4, respectively. The duration of the indicator field was set at 20 msec. and the masking field at 12 msec. Data were collected in four 24-trial blocks per session for the SL condition, three masking and one No-Mask Control, for 8 sessions. Data were collected in six 12-trial blocks per session, two mask delays for each of the three 12-L conditions (the No-Mask Control was considered a mask delay) over 24 sessions, for a total of 32 sessions. The assignment of conditions to sessions was counterbalanced within and partially counterbalanced across *S*s with the restrictions that no session contained the same combination of mask delay and condition twice, and that all combinations were assigned once before any were assigned a second time, and so on, for all sessions. This provided a number of counterbalanced points in the experiment where adjustments could be made in the duration of Field 1 to control for practice effects lowering *S*'s threshold. This had to be done an average of once for each *S* and condition.

Each *S* participated in three preexperimental sessions for purposes of setting thresholds and obtaining practice in all experimental conditions. Prior to each experimental session, *S* dark adapted for 5 min., after which they received 12 warm-up trials without a mask. The order of trials within a block was randomized by shuffling the stimulus cards. For the 12-L sessions, *S* first reported the letter the indicator was pointed at (first position), and then the letter directly opposite the indicated letter (second position), followed by the clock position of the indicator. This procedure, used by Eriksen and Collins (1969), encourages *S* not to move his point of fixation toward the indicated letter in the LE-FI condition and also provides a measure of any potential performance trade off between the first and second reported positions across the 12-L conditions. For the SL sessions, *S* simply reported the letter presented and its clock position. The *S* was instructed to respond on every trial, with a guess if necessary. All trials were self-initiated by *S* after *E* gave the ready signal, and *S* had the fixation cross in good focus. The *S*s were given feedback of the number of correct responses at the end of each block. All responses were recorded by *E*.

## RESULTS

The masking functions for all conditions are presented in Fig. 1, including the appropriate No-Mask Controls. A repeated-measures analysis of variance for the three 12-L first-position functions (12-L Condition × SOA × *S*s) indicated significance for the main effects of SOA, $F$ (5, 20) = 68.27, $p < .01$, and 12-L conditions, $F$ (2, 8) = 4.93, $p < .05$, and the SOA × 12-L Condition interaction, $F$ (10, 40) = 3.15, $p < .01$. A Newman-Keuls procedure for testing ordered pairs of means at each SOA indicated that the LE-FI function was significantly greater ($p < .05$) than the LE-SI function at SOA = 100, 150, 200, 250 msec., and significantly greater than the HE-SI function at SOA = 100, 150, and 200 msec. The LE-SI and HE-SI functions were not significantly different at any SOA by this test (although very close to significance at SOA = 150 and 250 msec.). A Dunnett's test (Winer, 1962) for comparing all means with a control indicated that the LE-FI function asymptoted to its No-Mask Control at SOA = 150 msec. as compared to SOA = 250 msec. for both the HE-SI and LE-SI functions. It is thus evident that an indicator presented 150 msec. prior to a 12-L display shortened the masking range for reporting the first position by approximately 100 msec. as compared to an equal or even higher energy 12-L display presented with a simultaneous indicator.

Unlike the results obtained by Eriksen and Collins (1969), an indicator preceeding the 12-L display by 150 msec. led to a nonsignificant increase in the No-Mask Control performance level. Increasing the energy of the display in the HE-SI condition resulted in a No-Mask Control performance level equal to that obtained in

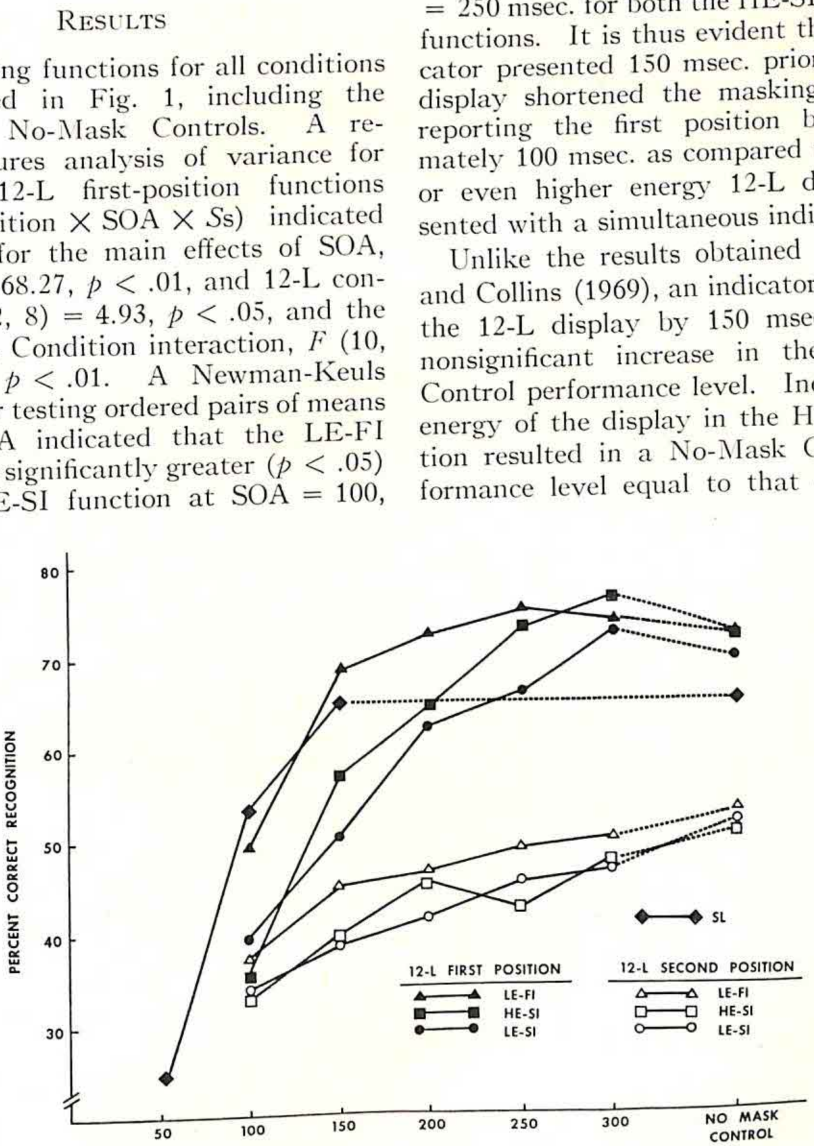


FIG. 1. Percent correct recognition (uncorrected for chance guessing) as a function of test form-pattern mask SOA for the SL function and the six 12-L functions.

the LE-FI condition (and also nonsignificantly greater than the LE-SI condition) but did not lead to a reduction in the masking range.

An analysis of variance for the SL function (SOA × *S*s) indicated significance for SOA, *F* (3.12) = 80.22, $p < .01$. Dunnett's test indicated this function asymptoted to its No-Mask Control at SOA = 150 msec., as did the LE-FI first-position function.

An analysis of variance for the three 12-L second-position functions in Fig. 1 indicated significance for SOA, *F* (5, 20) = 40.85, $p < .01$, and 12-L condition, *F* (2, 8) = 4.47, $p < .05$. The LE-FI condition was significantly better, overall, than both the LE-SI and HE-SI conditions. Dunnett's test indicated that the LE-FI function asymptoted at SOA = 250 msec., as compared to SOA = 300 msec. for the LE-SI function. The HE-SI function was less consistent with SOA = 200 msec., being nonsignificantly different from the No-Mask Control, while SOA = 250 msec. was significantly different. These results indicate that the shorter masking range obtained for the LE-FI first-position function was not a result of a performance trade off with the LE-FI second position since the function for the latter was equal to or greater than the other two second-position functions at all SOAs.

## DISCUSSION

The equivalence of masking range for the SL and 12-L LE-FI functions suggests that the icon formation time for a 12-L visual display occurs as rapidly as that of an SL display. The phenomenal report by *S*s also supports this conclusion. At short SOAs (e.g., 150 msec.), *S*s reported that an 12-L display appeared as fully developed as an SL display, but when presented with a simultaneous indicator there often wasn't enough time to both locate and fully encode the target letter. Apparently, a forward indicator provides *S* with additional time for this process. It would also appear that embedding the target letter in an array of letters (as in the 12-L conditions) did not significantly impede processing, as compared to an SL display, if *S* knew prior to the display's onset where to direct selective attention. An indicator was not used for the SL condition since Spencer (in press) found no differential masking range for an SL presented either near the fixation point without positional uncertainty or on an imaginary 2° circle with positional uncertainty. The SL function was included to represent the minimum masking range obtainable for the performance levels employed in the present experiment, although the SL function may have had a slightly shorter range if its No-Mask Control performance had been equal to that of the 12-L conditions. In either case, the forward indicator was very efficient in reducing the 12-L masking range as compared to a simultaneous indicator.

An attempt to account for the shorter masking range for the first position with a forward indicator as due to anticipatory eye movements would appear to be inconsistent with the second-position results. If *S* would or could move his point of fixation toward the expected location of the first-position letter as a means of both improving its clarity and reducing the search time (it would seem unlikely or difficult in 150 msec., and *S*s were instructed not to change their point of fixation), the second-position letter would be in a foveal location of even poorer clarity and recognition accuracy would be expected to suffer. However, the performance for the second position with a forward indicator was equal to or greater than that of both second-position indicator simultaneous functions at all SOAs, suggesting *S*s did not change their point of fixation. Nevertheless, a change in eye fixation with a forward indicator cannot be completely ruled out since the resulting poorer clarity of the second position could have been compensated by a higher quality stage of iconic storage concommittant with earlier encoding, as discussed below. Information concerning the relative trade off in performance between these two factors, and/or objective measures of eye movements, is required. But the possibility of altered fixation is even less convincing when one considers the relatively small improvement in the No-Mask Control performance level for the first position with a forward indicator.

A model more consistent with the data (and similar to that discussed by Eriksen and Collins, 1969, and Eriksen and Hoffman, 1972) is one in which a forward indicator directs selective attention (without eye movement) toward the first position, thereby reducing or eliminating search time for the first letter with a consequent reduction in the LE-FI masking range. This should also allow encoding to proceed off an icon of higher quality since the location of the critical item would occur at an earlier stage of iconic storage. Selective attention then proceeds to locate and process the second letter which, in the case of a forward indicator, would also occur sooner and at an earlier and presumably higher quality stage of iconic storage. This model would account for the generally greater masking range for the second position as compared to the first position and the shorter range for the second position with a forward indicator as compared to the second position with a simultaneous indicator (assuming the "premature" asymptote of the HE-SI second-position function at SOA = 200 msec. was due to sampling error). However, the nonsignificant improvement in the LE-FI No-Mask Control performance level for the first position, although in the predicted direction, was smaller than might be expected (Eriksen & Collins, 1969, obtained a significant improvement) since encoding off an icon of higher quality should improve performance. The obtaining of only a small improvement may have resulted from delaying the stimulus field 150 msec. after *S* initiated the trial in only the LE-FI condition. In this case, *S*'s focus may not have been optimal when the stimulus field was presented, reducing the advantage of a forward indicator with respect to recognition accuracy. The same explanation would apply to the lack of a significant improvement in the LE-FI second-position No-Mask Control performance level. However, the model would still predict significant reductions in masking ranges for both first and second positions with a forward indicator, as were obtained.

The present model proposes that encoding of two letters from iconic storage occurred in a serial (or at least nonparallel) manner. However, it should be noted that *S*s were instructed to give priority to the first-position letter as this was of primary interest. It may also be possible for *S*s to more equally divide their attention, and hence their performance, between the first and second positions when instructed to do so. This may or may not result in the same overall average performance level obtained in the present study, but a forward indicator would still be expected to provide an overall shorter masking range.

## REFERENCES

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell Systems Technical Journal*, 1961, **40**, 309–328.

Eriksen, C. W. Temporal luminance summation effects in backward and forward masking. *Perception & Psychophysics*, 1966, **1**, 87–92.

Eriksen, C. W., & Collins, J. T. Temporal course of selective attention. *Journal of Experimental Psychology*, 1969, **80**, 254–261.

Eriksen, C. W., & Hoffman, J. E. Some characteristics of selective attention in visual perception determined by vocal reaction time. *Perception & Psychophysics*, 1972, in press.

Spencer, T. J. Some effects of different masking stimuli on iconic storage. *Journal of Experimental Psychology*, 1969, **81**, 132–140.

Spencer, T. J. Backward masking as a function of stimulus energy and positioned uncertainty. *American Journal of Psychology*, 1972, in press.

Spencer, T. J., & Shuntich, R. Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, 1970, **85**, 198–203.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received August 21, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 302–308

# MEMORY IMPAIRMENT IN THE AGED: STORAGE VERSUS RETRIEVAL DEFICIT[1]

DAVID A. DRACHMAN[2] AND JANET LEAVITT

*Northwestern University*

A group of aged *S*s was compared with young control *S*s to elucidate the nature of the memory impairment occurring with normal aging. Tests of immediate memory span, supra-span digit storage, cued and uncued free recall, and retrieval from old storage were given to all *S*s. Aged *S*s showed no deficit in immediate memory span and retrieval from old storage, but showed highly significant impairment of learning of supra-span digits and verbal items. First-letter cuing provided an advantage over uncued free recall both for young and aged *S*s, but the cued recall advantage (CRA) was no greater for the aged *S*s. These findings do not support the view that a deficit of retrieval is the primary cause of the memory impairment of the aged and suggest that a disorder of storage is responsible.

The complaint of impairment of memory is well known to occur in a large proportion of aged individuals. Many studies have documented the differences in the ability of young and aged *S*s to learn new information under a variety of experimental circumstances (Bromley, 1958; Caird, 1966; Canestrari, 1968, pp. 169–188; Gilbert & Levee, 1971; Korchin & Basowitz, 1957; Kral, 1966). In the past, most such studies have employed the technique of free recall, in which *S* was required to learn new information (word lists, paragraphs, etc.); more recently, recall-recognition and cued–uncued free recall tests have been used in an effort to determine whether the deficit in memory in the aged is due to impairment of the hypothesized mechanisms of storage or retrieval. Thus, Schonfield (1965, 1967) and Schonfield & Robertson (1966) demonstrated that aged *S*s performed less well than young *S*s on free recall of a word list but did as well as the young *S*s on a recognition test. Laurence (1967) also found that aged *S*s did poorly on a free recall task but improved their performance when given category cues at the time of recall, again approximating the scores of young *S*s. Both authors concluded that these results showed the major memory deficit in the aged to be due to impairment of retrieval, while storage was essentially normal. Their reasoning, in accord with the views of Tulving (e.g., Tulving & Pearlstone, 1966) was that some information that had been stored was "available," but not "accessible," and that the large discrepancy between available and accessible information in the aged was due to some impairment of the "retrieval process" rather than to impaired storage.

McNulty and Caird (1966) have criticized Schonfield's (1966) conclusion on the grounds that partial learning of to-be-remembered (TBR) items is sufficient to make a correct choice in a recognition test and that the recognition scores of aged *S*s were "artificially inflated" by use of this technique. McNulty and Caird concluded that a deficit of storage, rather than of retrieval, might have produced the same findings. They did not, however, consider the implications of cued recall experiments in the evaluations of memory impairment in the aged.

To obtain more direct data on the question of storage and retrieval deficits in the aged, and to cast some light on the role of cuing in these processes, a series of experi-

---

[1] This research was supported in part by the Sterling Morton Research Fund. The facilities of the Northwestern University Neurological laboratory were used. The authors wish to express their appreciation to M. W. Laurence for generously providing the lists used in her (1967) study for our analysis and to thank Rita Dreiblatt, who aided in the collection of the data.

[2] Requests for reprints should be sent to David A. Drachman, Department of Neurology, Northwestern University—McGaw Medical Center, 303 East Chicago Avenue, Chicago, Illinois 60611.

ments were conducted on groups of young and aged *S*s.

## METHOD

### *Subjects*

*Experimental* (*aged*).—Twenty-four male and female adults aged 59 to 89 yr., with a mean age of 66.8 yr. served as aged *S*s. All had at least the equivalent of 4 yr. of high school and were well and active in the community with no neurological or psychiatric disease. They were recruited from sources such as old-age clubs, hospital volunteer corps, word-of-mouth, and referrals.

*Normal controls* (*young*).—Twenty-four male and female medical and nursing students and college undergraduates served as normal control *S*s. Their ages varied from 18 to 26, with a mean age of 21.8.

### *Procedure and Test Battery*

The *S*s were tested individually in one experimental session lasting approximately 2 hr. Testing was carried out in a sound-resistant room with *S* facing *E*. The following tests were presented in a fixed order to provide an appropriate control situation for between-group comparison of *S*s:

*Wechsler Adult Intelligence Scale* (*WAIS*).—This test was administered to all *S*s to provide a standardized measure of general cognitive functioning and to aid in habituation to the experimental situation, particularly for aged *S*s.

*Single-trial span.*—This is similar to the standard digit-span test used in both the WAIS and the Wechsler Memory Scale. Random sequences of nonrepeating digits were presented by means of a tape recorder at the rate of one digit per second. The *S*s were instructed to repeat orally in the order of presentation all the digits they remembered. Two different sequences of each length were presented starting with four digits and increasing until *S* failed to recall either sequence perfectly. The longest sequence perfectly recalled at least once was considered *S*'s digit span. The data were scored in the standard manner with recalled sequences counted as correct only if all digits were recalled and appeared in their appropriate serial locations (Wechsler, 1955, p. 40).

*Supra-span digit storage.*—This test consisted of the repeated presentation of each of five sequences of 15 digits. Digit sequences were constructed from random tables using the numbers 1–9. In each sequence no digit occurred more than twice, adjacent digits could not be identical, and no part of a sequence could be a generally familiar grouping (e.g., 1492). Digit sequences were presented by means of a tape recorder at the rate of 1 digit per second. Presentation of each sequence (training trials) alternated with attempts by *S* to reproduce the sequence (testing trials) for eight trials, or until a criterion of one perfect recall was achieved. The next sequence was then initiated and the same procedure was followed until each of the five sequences had been tested. The *S*s were instructed to repeat orally in the order of presentation all the digits they remembered even if uncertain of the correctness of the response. For each testing trial the number of correct responses was recorded. Scoring was designed to recognize as many correctly recalled items as possible, while minimizing scores for chance responses by employing the following rules: (*a*) The first and last items in a series, if correctly recalled in that position, were scored as correct; (*b*) any correct items adjacent to correct first or last items were scored as correct; (*c*) all groups of three or more items whose order corresponded to a similar sequence in any location in the digit series were scored as correct. The advantages of this scoring method are demonstrated mathematically elsewhere (Drachman & Zaks, 1967). Mean items correct for each trial from one to eight were calculated for the five digit series. Learning curves were derived from the data, and the slope and area under the curve mathematically determined. This mathematical treatment (*a*) provided an accurate measure of the rate of new learning, (*b*) permitted the computation of comparable scores for *S*s who reached criterion in fewer than eight trials, and those who did not, and (*c*) compensated (area measure) for the possibility of relatively large variations in first-trial recall.

*Cued and uncued free recall.*—In this test, the critical experimental manipulation was the presence or absence of cuing during recall. A multitrial free recall paradigm was employed with the initial letters of the list words used as cues. The *S*s were tested for both cued and uncued recall, and for each *S* a different list was used for each recall condition. Recall was written and *S*-paced.

Two equivalent lists of 35 words each (A and B), arranged nonalphabetically, were constructed from 70 common English words selected at random from Level AA of the Thorndike-Lorge word count. An invariant word order was used to assure that all *S*s would readily recognize that the list was being repeated and to minimize the possible confusion that varying the word order might produce particularly in aged *S*s. Lists and recall order (cued first or cued second) were counterbalanced for four groups each of young and aged *S*s.

Standard free recall instructions were given for uncued recall. In addition, for cued recall, *S* was told prior to list presentation that after he had seen all the words he would be given an answer sheet containing the initial letters of all words in the order of presentation. All *S*s were shown a sample answer sheet, which was removed before testing was initiated.

Simultaneous auditory and visual presentation of words was employed to control for modality effects. Words were presented by the use of a Sony Model 230 tape recorder synchronized to operate a Kodak Carousel Model 800 projector. Each word was exposed individually for 2.5 sec. on a rear-projection screen. Immediately after the last word appeared, *S* was given an answer sheet appropriate to the recall condition and told to write all the words he remembered. For cued recall, he was told to write TBR words after the appropriate initial letters. When *S* indicated that he was finished writing, another trial was initiated. Three consecutive trials of "study-test" in one recall condition (cued or uncued) were followed by three consecutive trials in the other.

The four response measures considered were (*a*) number of words correctly recalled in each recall condition, (*b*) "cued recall advantage" (CRA) calculated as correct recall of

$$\frac{\text{cued-uncued}}{\text{uncued}},$$

(*c*) learning curve slopes, and (*d*) intrusions.

*Retrieval by category.*—In this test, *S* was given the name of a familiar category of nouns and instructed to say as many words that fit the category as rapidly as he could, until *E* told him to stop. An example was given to insure that *S* understood the directions. Five categories were used for retrieval: (*a*) fruits, (*b*) flowers, (*c*) girls' names, (*d*) vegetables, and (*e*) animals. The order of the categories was the same for all *S*s, and spoken recall was recorded manually by *E*.

A stopwatch in full view of each *S* was used to time production for the first 20 sec. and for a total of 60 sec. A new category was introduced as soon as the previous one was completed. Scoring was the number of appropriate nonrepeated responses emitted in each time period.

## RESULTS

*Wechsler Adult Intelligence Scale.*—WAIS scores for young and aged *S*s showed no significant differences. The mean Full Scale IQ was 117.8 for young *S*s, 117.2 for aged *S*s, $t$ (44) = .23, *ns*. The mean Verbal IQ was 119.6 for young *S*s, 120.8 for aged *S*s, $t$ (45) = .42, *ns*; the mean Performance IQ was 113 for young *S*s, 110.2 for aged *S*s, $t$ (44) = 1.11, *ns*. These data support the assumption that young and aged *S*s were comparable in general cognitive ability.

*Single-trial digit span.*—Digit spans were not significantly different for young (7.2) and aged (6.3) *S*s. A $t$ test comparing mean digit span scores of young and aged *S*s yielded $t$ (46) = 1.96, *ns*. This is consistent with other experimental studies which report no significant age-related differences in immediate memory span (Bromley, 1958; Drachman & Hughes, 1971; Kriauciunas, 1968).

*Supra-span digit storage.*—Aged *S*s were markedly inferior to young *S*s in the rate of learning of supra-span digits. The mean slopes of 1.86 for young and .78 for aged *S*s indicate that the rate is considerably depressed in the aged. The $t$ test comparing mean slopes between young and aged *S*s yielded $t$ (46) = 3.57, $p < .001$. Figure 1 graphically illustrates this highly significant difference.

The mean area under the learning curve for young *S*s was 75.40 compared with a mean area of 46.53 for aged *S*s. Comparison of areas showed a highly significant difference between young and aged *S*s, yielding $t$ (46) = 3.83, $p < .001$. Pearson product-moment correlation coefficients comparing slopes with area scores were .96 for young *S*s, .71 for aged *S*s, and .95 when all *S*s were pooled into a single group. This indicated that both in the rate of accumulation of information (slopes) and in total information recalled (areas), young *S*s were highly superior to aged *S*s and that the two measures were closely correlated.

*Cued-uncued free recall.*—The total numbers of words recalled over all trials for young and aged *S*s, both in the uncued and the cued conditions were compared. In the uncued mode young *S*s recalled a mean of 58.17 words, aged *S*s 35.92; in the cued mode young *S*s recalled a mean of 63.50 words, aged *S*s 38.33. *T* tests showed that the recall of young *S*s was significantly better than that of aged *S*s, both in the uncued and cued modes, $t$ (46) = 6.43 and 6.36, respectively, $p < .001$.

A CRA score was calculated using the formula:

$$\text{CRA} = \frac{\text{cued-uncued}}{\text{uncued}}$$

to test the hypothesis that cuing aids recall. Mean CRA scores over the three trials for all *S*s were pooled and a $t$ test analysis performed to determine whether a cuing advantage does indeed occur. This analysis yielded a mean CRA of .13, $t$ (46) = 2.47, $p < .02$, thus showing a definite advantage for the cued condition. The CRAs for young *S*s were then compared with those for aged *S*s, to see if cuing provided a greater relative improvement in recall for aged *S*s. The mean CRA for both young and aged *S*s was .13. An analysis

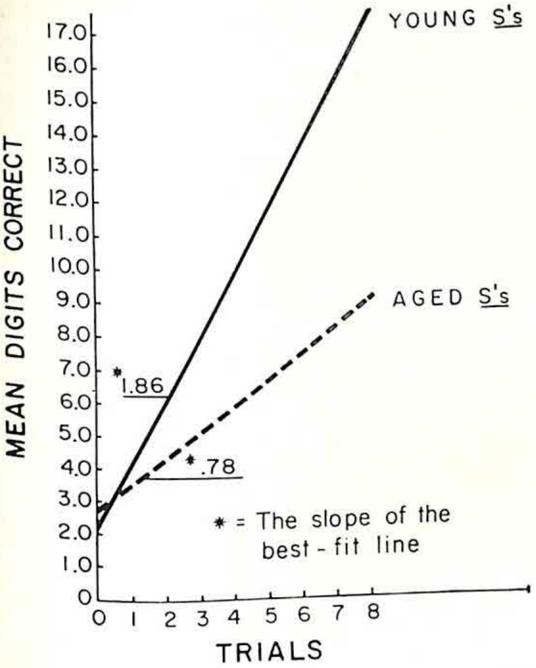


FIG. 1. Mean learning curves of young and aged *S*s for supra-span digits plotted by mean slope and *y* intercept.

of variance showed no significant difference between CRAs for young and aged *S*s. The age variable yielded $F\ (1, 44) = .0008$, *ns*. This indicated that cuing did not provide a relatively greater recall advantage for aged *S*s.

Slopes of the learning curves were calculated for verbal TBR items over three trials, both in the uncued and cued modes. In the uncued mode, the mean slope for young *S*s was 6.9 compared with 4.2 for aged *S*s; in the cued mode the mean slope for young *S*s was 7.7 compared with 4.7 for aged *S*s (Fig. 2). *T* tests again showed that young *S*s performed significantly better than aged *S*s, both in the uncued and cued modes, $t\ (46) = 4.69$ and 5.15, respectively, $p < .001$. This indicated that the rate of accumulation of verbal information was better for young than for aged *S*s and is consistent with the data for digit series.

Extralist intrusions and list intrusions for the second recall condition were determined for young and aged *S*s. None of the within-*S* or between-group analyses for list intrusions revealed significant differences. The only significant finding from all analyses was the increased number of extralist intrusions in the cued condition, both for young and aged *S*s. In the uncued mode, the mean number of extralist intrusions was 1.21 for young *S*s, 3.33 for aged *S*s; in the cued mode, 3.46 for young *S*s, 5.88 for aged *S*s. The differences between cued and uncued extralist intrusions was significant both for young *S*s and for aged *S*s. Statistical analyses yielded $t\ (23) = 3.73$, $p < .01$, for young *S*s, and $t\ (23) = 2.60$, $p < .02$, for aged *S*s. The differences between young and aged *S*s did not reach significance in either recall mode.

A *percentage learning rate* (PLR) was calculated comparing the rate of accumulation of new information for aged *S*s compared with young *S*s:

$$\text{PLR} = \frac{\text{mean slope aged } Ss \times 100}{\text{mean slope young } Ss}.$$

The PLR thus provided an index of the "efficiency" of aged *S*s in learning supra-span digit sequences and supra-span word lists. Percentage learning rate for digit sequences was 42% and PLR for word lists was 61%, suggesting relatively greater im-

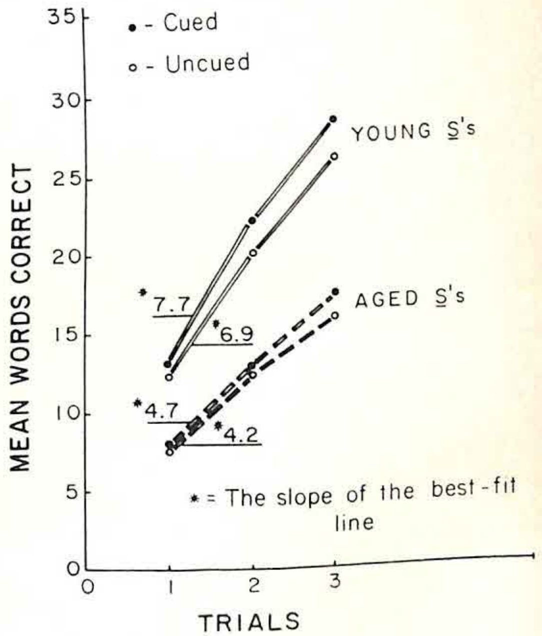


FIG. 2. Mean number of words correctly recalled over three trials for young and aged *S*s with cued and uncued recall.

pairment in learning supra-span digit series than supra-span word lists for aged *S*s.

*Retrieval by category.*—There was no significant difference in the amount of information retrieved by young and aged *S*s from familiar categories. Mean scores for the first 20 sec. of retrieval were 45.2 for young *S*s, 41.1 for aged *S*s; total retrieval scores for 1 min. were 80.8 for young *S*s, 76.1 for aged *S*s. *T* tests comparing performance of young and aged *S*s for retrieval during the first 20 sec. and for the total 1 min. were $t$ (46) = 1.57 and .98, respectively, *ns*.

## Discussion

The series of memory studies carried out in this investigation are useful in attempting to evaluate the relative importance of storage and retrieval deficits in the memory impairment of the aged. First, the possibility was considered that some alteration of alertness, attention, or perception might account for the memory impairment noted. Direct observation by *E* indicated that the aged *S*s were alert, paid close attention during testing, and were highly motivated to perform well; further, their unimpaired performance on the digit span (Pope & Scott, 1967, p. 74) and retrieval from old storage tests, which were presented toward the beginning and end of the experimental sessions, respectively, suggested that attentiveness was sustained throughout. Inglis (1964) has also concluded that the memory impairment of the aged cannot be attributed to a defect of attention, perception, or motivation, in accord with the present findings.

Second, the impaired learning of supra-span digit series and word lists indicates a definite memory impairment; it may be argued that this is the result of either a deficiency of storage, or a deficiency of retrieval, or both. Closer scrutiny of these data may provide some useful, if indirect, insights into the relative impairment of storage or retrieval in the aged, however. Digits are exceedingly familiar to all *S*s, and the entire set (1–9) is instantly available. For this reason, learning and recall of supra-span digit series probably depend more on storage (i.e., forming order linkages and organization of groups) than on retrieval, or search, for the individual items (Murdock, 1968). By comparison, the words in a list are derived from an immeasurably larger vocabulary; learning and recall depend not only on storage, but also, to a greater extent than with digits, upon efficient search for, and retrieval of, the items. If this is so, one would anticipate that *S*s with a disorder of retrieval would show relatively greater impairment in a word-learning than in a digit-learning test. In our studies, this was not the case: aged *S*s achieved a lower PLR for digits than for words. This comparison thus failed to support the hypothesis that the major cause of the memory impairment was a retrieval deficit.

First-letter cuing of word lists at recall provided an advantage for all *S*s over the uncued mode, but cuing failed to provide a relatively or absolutely greater advantage for the aged *S*s. A current view suggests that cuing should provide a greater recall advantage for *S*s who have some impairment of retrieval; by this interpretation, the lack of a significantly greater CRA for the aged *S*s might indicate that retrieval was relatively unimpaired in the aged.

The study of retrieval by category is of special interest in that it revealed no significant impairment in the aged *S*s. This is the only direct test of retrieval in the battery of studies, and intactness of performance is a strong indication that retrieval, at least from old memory stores, is unimpaired. It is unclear at present whether the operations involved in retrieval by category are the same as those required for recall of recently stored information. The present investigation does not resolve this question, and further studies are needed to clarify this point (Tulving, 1970, pp. 7–9).

The data obtained in this series of experiments thus provide evidence for an impairment of some, but not all, memory functions in the aged. Immediate memory span is unimpaired; the learning of supra-span lists shows a major deficit, however. Retrieval by category, a direct test of retrieval, shows no impairment in the aged; similarly, indirect measures of retrieval fail to provide evidence for a disorder of retrieval as the prime cause of the observed memory impairment. Our results thus suggest that an impairment of memory storage was most likely the primary cause of the deficit.

How then can one explain the results of Schonfield (1965, 1967) and of Laurence (1967), who concluded that retrieval, and not storage, was deficient in the aged? Schonfield's conclusion that the memory impairment of the aged is due to a retrieval deficit, based on his observation that aged *S*s performed as well as young *S*s on recognition tests while significantly impaired on free recall tests, is probably not justifiable. We would agree that recognition tests reduce the required amount of retrieval to a marked degree, but they may also permit partially learned items to be fully recalled, increasing the apparent amount of remembered information when much incomplete learning has occurred. McNulty and Caird (1966) have similarly objected that in a recognition test "only partial input images (two or three letters of an item . . . ) need to be stored in order to recognize the input materials completely [p. 229]." Thus, a recognition test is not necessarily an exact measure of storage since with either an impairment of storage or of retrieval it may show less of a deficit than a free recall test. With a storage impairment, recognition scores may be inflated by crediting partially learned items; with a retrieval deficit, it may be aided by the reduction of the need for retrieval.

Laurence (1967), however, did not use recall-recognition tests, but instead compared the performance of young and aged *S*s on free recall, with and without cues. In the uncued mode, aged *S*s recalled less than young *S*s; with category cues at recall both groups performed equally well. Laurence's conclusion, like Schonfield's (1966), was that providing a cue at recall was useful in overcoming a retrieval deficit in aged *S*s.

Careful analysis of Laurence's (1967) study, including review of the actual word lists used (M. W. Laurence, personal communication, November 1971), suggests that the cued mode may have functioned in part like a recognition test. The cuing categories were all common ones (flowers, countries, birds, etc.); almost half of the items were common enough so that according to the norms of Battig and Montague (1969), 20% to 37% of untrained *S*s would emit them spontaneously if given only the category. Using these norms, slightly over seven of the actual list words, on the average, would be available to *S*s simply as items generated by the category cue. These words would then be available for a recognition-like discrimination, once again permitting partially learned items to be identified and favoring aged *S*s who may have fewer items fully learned. Since aged *S*s improved their recall by approximately four words with the category cues, the recognition-like effect postulated above is of an appropriate magnitude to account for the improvement. Using TBR items from known lists, Slamecka (1967) has documented this phenomenon, showing that normal *S*s performed as well on free recall as on recognition tests if the TBR items are from a single very familiar category. Initial letter cues would not produce a recognition-like situation, since they cue a much larger potential vocabulary, and would be unlikely to generate any of the list words by chance.

Although initially we had assumed that the difference between cued and uncued recall was a "retrieval deficit," where the CRA represented stored but inaccessible information, other considerations have suggested that this is an oversimplification. The correctness of this concept depends on an implicit acceptance of the concept that information is stored in an "all-or-none" fashion. If storage is all or none, then information that has been stored should be retrievable unless the retrieval mechanism is inefficient; the greater the deficiency of retrieval, the greater the benefit from cuing at recall.

It is far more likely, however, that information may be learned with variations in strength of storage (Wickelgren, 1970, pp. 65–102). If this is so, then a cue presented at the time of recall could be of benefit with either deficiencies of retrieval or of storage. An explanation for this depends on the concept of a "threshold" level for recall of stored items. In a list-learning experiment, a "completely learned" item will have achieved sufficient signal strength so that during recall it will exceed a threshold level and be spontaneously emitted by *S*. A "partially learned" item would not have sufficient signal strength to reach threshold and be spontaneously emitted; yet it may have a signal strength greater than zero. If retrieval is impaired, the threshold required to achieve recall may be raised above normal; if storage is impaired, even the normal threshold may not be reached because of weak signal strength. In either case, a second means of retrieval, such as the use of a category cue at recall, could boost the signal strength of the TBR item (along with other items in the same category). The "summation" of signal strengths for the item derived from the category cue in addition to the word list may then exceed threshold; if so, it will be recalled. Inevitably, the use of category cues at recall must increase the number of intrusions, since the signal strengths of all items in the same category are increased and the signal/noise ratio is not improved for the TBR item. This, indeed, has been the observation in a number of studies of cuing including the present one; more correct items are recalled, along with more intrusions from the cuing category (Tulving & Osler, 1968; Tulving & Pearlstone, 1966).

This view of the mechanism by which cues provided at recall are effective implies that a cuing advantage found at recall would be ambiguous; alone, it would not discriminate between a storage or a retrieval deficit in the aged. Further, it predicts that the stronger the cue (of which a recognition task would be the most powerful), the greater the CRA for *S*s with impaired memory due to deficits either of storage or retrieval.

If these considerations are correct, then what inferences about storage and retrieval mechanisms can be drawn from human memory studies? Only retrieval by category provides a direct measure of either function and is probably an almost pure test of retrieval from old store when the stores themselves are not impaired (as in certain aphasias). Recall-recognition and cued-uncued learning tests may provide some indirect measures of storage and retrieval, but they rely on certain assumptions which require further verification. Viewing our studies of aged *S*s in this light, we have been able to find no evidence for a retrieval deficit as the major cause of the observed memory impairment; a disorder of storage seems more likely at present.

## REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

BROMLEY, D. B. Some effects of age in short-term learning and remembering. *Journal of Gerontology*, 1958, **18**, 398–406.

CAIRD, W. K. Aging and short-term memory. *Journal of Gerontology*, 1966, **21**, 295–299.

CANESTRARI, R. J., JR. Age changes in acquisition. In G. A. Talland (Ed.), *Human aging and behavior*. New York and London: Academic Press, 1968.

DRACHMAN, D. A., & HUGHES, J. R. Memory and the hippocampal complexes. III. Aging and temporal EEG abnormalities. *Neurology*, 1971, **21**, 1–14.

DRACHMAN, D. A., & ZAKS, M. S. The memory cliff beyond span in immediate recall. *Psychological Reports*, 1967, **21**, 105–112.

GILBERT, J. G., & LEVEE, R. F. Patterns of declining memory. *Journal of Gerontology*, 1971, **26**, 70–75.

INGLIS, J. Influence of motivation, perception and attention on age-related changes in short-term memory. *Nature*, 1964, **204**, 103–104.

KORCHIN, S. J., & BASOWITZ, H. Age differences in verbal learning. *Journal of Abnormal and Social Psychology*, 1957, **54**, 64–69.

KRAL, V. A. Memory loss in the aged. *Diseases of the Nervous System*, 1966, **27**, 51–54.

KRIAUCIUNAS, R. The relationship of age and retention-interval activity in short-term memory. *Journal of Gerontology*, 1968, **23**, 169–173.

LAURENCE, M. W. Memory loss with age: A test of two strategies for its retardation. *Psychonomic Science*, 1967, **9**, 209–210.

MCNULTY, J. A., & CAIRD, W. Memory loss with age: Retrieval or storage? *Psychological Reports*, 1966, **19**, 229–230.

MURDOCK, B. B., JR. Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology*, 1968, **77**, 79–86.

POPE, B., & SCOTT, W. M. *Psychological diagnosis in clincial practice*. New York: Oxford University Press, 1967.

SCHONFIELD, D. Memory changes with age. *Nature*, 1965, **208**, 918.

SCHONFIELD, D. Memory loss with age: Acquisition and retrieval. *Psychological Reports*, 1967, **20**, 223–226.

SCHONFIELD, D., & ROBERTSON, B. Memory storage and aging. *Canadian Journal of Psychology*, 1966, **20**, 228–236.

SLAMECKA, N. J. Recall and recognition in list-discrimination tasks as a function of the number of alternatives. *Journal of Experimental Psychology*, 1967, **74**, 187–192.

TULVING, E. Short- and long-term memory: Different retrieval mechanisms. In K. H. Pribram & D. E. Broadbent (Eds.), *Biology of memory*. New York and London: Academic Press, 1970.

TULVING, E., & OSLER, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593–601.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

WICKELGREN, W. A. Multitrace strength theory In D. A. Norman (Ed.), *Models of human memory*. New York and London: Academic Press, 1970

(Received August 23 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 309–319

# IMPLICIT ASSOCIATIVE RESPONSES IN VERBAL DISCRIMINATION ACQUISITION[1]

N. JACK KANAK,[2] LAWRENCE E. COLE,[3] AND ED ECKERT

*University of Oklahoma*

Predictions from the frequency theory of verbal discrimination learning for interitem manipulations of implicit associative responses were examined in four experiments varying methodological factors. In general, the predictions with respect to lists involving associatively related right items were supported only when *S*s were explicitly instructed regarding the associative relationships. In lists involving associatively related wrong items, the predicted inferiority relative to the control condition was obtained only in Exp. II. Both experimental and control lists were learned more readily under the study-test method as opposed to anticipation, with an initial study trial, with longer durations of presentation, and, in general, by females. The role assigned by the frequency theory was discussed for interitem, interpair, and intrapair manipulations of implicit associative responses.

The frequency theory of verbal discrimination (VD) learning (Ekstrand, Wallace, & Underwood, 1966) suggests that implicit associative responses (IARs, Bousfield, Whitmarsh, & Danick, 1958) may facilitate or interfere with the acquisition of VD lists employing associatively related words, depending upon whether the IAR increases or decreases the relative differential subjective frequency of occurrence of the wrong (W) or right (R) items of a given pair. When a list involves associatively related R items and unrelated W items (AR condition), acquisition is expected to be facilitated as a result of the frequency increment to R items via IARs, relative to a control (C) list of unrelated W and R items. Conversely, a list composed of associatively related W items and unrelated R items (AW condition) is assumed to reduce the differential frequency cue and should thus be more difficult than the C list. The above forms of manipulating IARs in VD lists will hereafter be inferred to as interitem manipulations or lists as distinguished from intrapair (i.e., the associated items are the components of a given W-R pair) and interpair (the associated items have W and R functions in different pairs) manipulations (e.g., Eberlein & Raskin, 1968; Fulkerson & Kausler, 1969). The present article reports the results of four experiments which tested the generality of predictions from frequency theory for interitem IAR lists across a number of methodological factors.

Ekstrand et al. (1966) found that an AR condition did result in fewer errors over four trials than the C condition, but the predicted inferiority of the AW condition was not obtained. Kanak, Cole, and Thornton (1970), in two experiments, failed to find the AR condition superior to the C condition, and the AW condition resulted in the predicted inferiority in only one of three comparisons on errors over the first four trials. The Ekstrand et al. and Kanak et al. studies differed in methodology in a large number of potentially important ways, including methods of presentation, sensory mode of presentation, the use of an initial guessing trial, list materials, and temporal rate of presentation. The series of experiments reported here examined these as well as two other methodological factors (instructions and sex of *S*s) in the AR, AW, and C conditions in an attempt to: (*a*) reconcile the discrepancies in the literature and (*b*) determine, as noted above, the generality of the predictions from frequency theory across methodological variables. Finally, since most

---

[1] The authors are grateful to Linda Scott, who collected the data of Exp. I. This research was supported in part by the Faculty Research Fund at the University of Oklahoma.

[2] Requests for reprints should be addressed to N. Jack Kanak, Department of Psychology, University of Oklahoma, Norman, Oklahoma 73069.

[3] Now at Arkansas Polytechnic College, Russellville, Arkansas.

methodological variables were varied factorially, the experiments also allow a systematic examination of the effects of these variables upon the VD learning of a standard list of unrelated items (i.e., the C condition).

## EXPERIMENT I

Experiment I varied method of presentation, i.e., anticipation (ANT) vs. study test (ST) in an attempt to determine whether method might interact with VD acquisition in IAR manipulations. Although a number of studies (e.g., Battig & Switalski, 1966; Fulkerson & Johnson, 1971; Ingison & Ekstrand, 1970; Schulz & Hopkins, 1968) have demonstrated superiority of the ST or "recall" method to ANT, no evidence has been presented to illustrate that method of presentation interacts with other VD task variables. Nevertheless, it is conceivable that IARs might more effectively operate under the ST method since the separation of learning from performance, or the avoidance of the alternation between storage and retrieval process associated with the ANT method, might facilitate the evocation of implicit associative responses. In addition, Exp. I varied the conduct of the initial trial (i.e., a guessing trial vs. no guessing). Frequency theory clearly predicts the superiority of no guessing or an initial study trial, and the available evidence (e.g., Dominowski, 1966) supports the theory when both items are presented in the feedback interval. To the extent that IAR activity is affected by *S*s immediate knowledge of correct items on the initial trial, no-guessing conditions could be more favorable to the IAR hypothesis. Finally, since list materials may be a particularly strong source of variance in recognition-discrimination learning (e.g., Kanak et al., 1970), both the original 20-pair list used by Ekstrand et al. (1966) and a list derived from their materials were used in Exp. I.

### *Method*

*Subjects.*—One hundred and twenty students enrolled in introductory psychology courses at the University of Oklahoma participated in the experiment in partial fulfillment of a course requirement. None of the *S*s had participated in prior verbal learning experiments. Treatment conditions were randomly assigned to *S*s upon their appearance at the laboratory.

*Materials.*—The C list of unrelated words was identical to that employed by Ekstrand et al. (1966) with one exception. The word SOFT, used by those authors as a W item, was eliminated since it bears a rather high associative relationship to LOUD, a word used elsewhere in their list as an R item. The word HERE was substituted for SOFT, and its associate THERE was used (when appropriate) on the AR and AW lists. Item function (i.e., W or R) of the words in the C list was counterbalanced such that one-half of the *S*s were given a list with W and R item functions identical to Ekstrand et al., while the remainder received a list with the above word functions reversed. This procedure was employed to yield greater generality of results and to match comparable counterbalancing procedures in the AR and AW lists.

One set (Set A) of AW and AR lists was identical to those employed by Ekstrand et al. (1966). A second set (Set B) of lists for these conditions was generated by selecting the remaining 10 R words of the original C list, not used (no asterisks) by Ekstrand et al. in their AR and AW conditions, and pairing these R items and their associates with the W items of the original C list. Therefore, all 20 original R words from the Ekstrand et al. C list were represented, with their associates in the AR and AW conditions, rather than only 10. Words functioning as associates were counterbalanced between the AR and AW conditions (as noted above) in both the original Ekstrand et al. list and the newly derived list.

Four random orders of serial position were employed for all list conditions. In addition, the left–right spatial position of R items was independently randomized for each serial order, with the restriction that successive occurrences of R items occupying the same spatial position not exceed three.

*Procedure.*—The lists were presented on a Lafayette memory drum for both the ANT and ST methods. All *S*s were tested individually. In the ANT method, a given pair appeared in horizontal alignment for 2 sec. during the anticipation interval and *S* pronounced aloud the item believed correct. The pair reappeared in the feedback interval, again for 2 sec., with the R item underlined. During the study trials of the ST procedure, the 20 pairs were exposed individually for 2 sec. each, with the correct item underlined. On test trials, the pairs were again exposed for 2 sec. each and *S* responded overtly as in the ANT method. The intertrial interval was 4 sec. for both methods of presentation.

Those *S*s assigned to the guessing (G) condition were instructed to guess on the first trial. This necessitated that *S*s learning by the ST procedure be given the first trial as a test trial, while ANT method *S*s were, of course, given the first trial as an anticipation trial. In the no-guessing (NG) condition, *S*s learning by the ANT method were given the initial trial as a study trial while ST *S*s, of course, also were given the initial trial as a study trial.

Standard instructions conforming to the basic requirements of the ANT and ST methods, as appropriate, were given to each *S*. Consistent with Ekstrand et al. (1966) and Kanak et al. (1970), no mention was made of the associated items in the AR and AW conditions. Practice was carried to a criterion of 2 successive perfect trials, or 15 trials, whichever occurred first.

### *Results*

The measures of learning included errors on Trials 1–4 for the no-guessing conditions, or on Trials 2–5 for the guessing conditions, and percent of total errors to the base of opportunities.[4] A preliminary analysis comparing List Sets A and B of the AR and AW conditions on each of the above measures indicated that neither the main effects for list sets or any of the interaction effects involving list sets with list conditions, method of presentation, or the guessing factor approached significance (all $p\text{s} < .10$). In addition, the two C lists appeared comparable in difficulty with the overall analysis of variance in each case yielding $F < 1$. List sets were therefore pooled within each condition and ignored for all subsequent analyses.

*Errors on Trials 1–4.*—The number of errors on Trials 1–4 (NG conditions) or 2–5 (G conditions) represents the most direct comparison with the data reported by Ekstrand et al. (1966). Only 5 *S*s of the total of 140 attained the two-perfect trial criterion in four or fewer trials, as defined above. For purposes of the analysis, these *S*s were assigned scores of zero errors on the postcriterion trials. List conditions yielded $F < 1$ with means and standard deviations, respectively, of 14.20, 8.53 (AR), 16.00, 9.89 (AW), and 16.47, 10.18 (C). Learning by the ST method ($\bar{X} = 11.38$, $SD = 7.65$) produced significantly fewer errors than the ANT method ($\bar{X} = 19.73$, $SD = 9.45$), $F\ (1, 108) = 27.44$, $p < .001$. The guessing variable was nonsignificant, $F\ (1, 108) = 1.15$, $p < .25$, as were all the interactions which typically produced $F < 1$.

*Total errors to the base of opportunities.*—Since perseveration errors may result in an overestimation of task difficulty when difficulty is measured by trials to criterion and since a substantial number of *S*s (10 in the C condition, 5 in the AR, and 6 in the AW) failed to achieve perfect performance within 15 trials, an analysis was conducted on the proportion of errors to the base of opportunities, i.e., total errors divided by the number of trials given multiplied by the number of pairs. For reasons noted earlier, this measure of task difficulty is preferred in all experiments to trials to criterion as a measure of difficulty, particularly when the relatively subtle effects associated with IARs are involved. List conditions again failed to produce any notable variability on the measure of errors to the base of opportunities, $F\ (2, 108) = 1.24$, $p < .25$. The means and standard deviations were

---

[4] Analyses were also conducted in all four experiments on trials to criterion, assigning a score of 17 to those *S*s who failed to achieve one perfect trial by Trial 15 or a score of 16 to those *S*s whose fifteenth trial was perfect. These analyses in all cases resulted in statistical conclusions consistent with the error analyses reported throughout the four experiments. Because errors represent a more sensitive measure of discrimination learning than trials to criterion, the present report omits the reporting of the trials to criterion analyses in all experiments.

.138, .064 (AR), .151, .066 (AW), and .158, .066 (C). The method of presentation variable (ANT $\bar{X} = .171$, ST $\bar{X} = .126$) was again significant, $F$ (1, 108) = 17.73, $p < .001$, and in addition, no guessing ($\bar{X} = .125$, $SD = .077$) produced a significantly lower error rate than the guessing condition ($\bar{X} = .172$, $SD = .056$), $F$ (1, 108) = 19.05, $p < .001$, in accordance with the frequency theory.

## EXPERIMENT II

The results of Exp. I failed to support the frequency theory with regard to the prediced role of IARs in the AR and AW conditions. Further, the data failed to replicate the findings of Ekstrand et al. (1966) with either the ST method or the original materials used by those authors.

The differences in error rates on the first four trials between the Ekstrand et al. (1966) experiment and Exp. I were large, with the rate being much higher in the present study. The authors reasoned that if *S*s were issued instructions pointing out the associative relationships in the AR and AW lists, *S*s should then use IARs in their learning strategy in a manner predicted from the theory, and, as a result, error rates should show a significant decrease as well. Experiment II, therefore, varied the instructional set given to *S*s in those two conditions, but using only the ANT method, an initial guessing trial, and the original lists employed by Ekstrand et al. Choice of the ANT method and a guessing trial was governed by the fact that these are the procedures most typical of previous VD research and by the lack of any interaction of method of presentation and guessing in Exp. I. Experiment II thus involved a factorial manipulation of instructions in the AR and AW groups, with the fifth group being an uninstructed control group.

### *Method and Procedure*

A total of 60 *S*s participated in the experiment, selected and assigned in the same manner as in Exp. I. The *S*s in the instructed (I) conditions were shown examples of the appropriate associative relationships from pairs of words not represented in the experimental list. In the AR-I condition, *S*s were told that each right item was an associate of another R item elsewhere in the list and that this information might prove to be helpful in learning the list. The *S*s in the AW-I condition were given similar instructions with regard to W items. Care was taken to make clear to *S*s that only two R items (or two W items) were related and that 10 such relationships could be found in the list.

No instructions beyond the standard instructions used in Exp. I were given to the AR-UI (uninstructed), AW-UI, and C condition groups. Twelve *S*s were assigned to each of the five groups in the same manner as in Exp. I. All other procedural details were identical to those of the first experiment.

### *Results*

*Errors on Trials 2–5.*—Three *S*s required only four trials to achieve criterion (one *S* in each of the C, AW-I, and AR-I groups), and these *S*s were recorded as having no errors on Trial 5 for the analyses of errors on Trials 2–5. A 2 × 2 analysis of variance indicated that the type of IAR list (i.e., AR and AW) and instruction main effects were both significant, $F$ (1, 44) = 13.17 and 6.98, respectively, $p$s $< .005$ and .025. The interaction was nonsignificant ($F < 1$). The mean errors for the IAR lists were 23.33 (AW) and 14.75 (AR), and the comparable measures for the instruction factor were 22.16 (NI) and 15.87 (I). The between-groups analysis of variance, including the C group, also resulted in a significant effect, $F$ (4, 55) = 4.86, $p < .01$. The means and standard deviations were 12.17, 8.62 (AR-I), 17.33, 8.69 (AR-NI), 17.92, 9.29 (C), 19.67, 6.83 (AW-I), and 27.00, 6.65 (AW-NI). It may be observed that the rank of each of the two AR and two AW groups, relative to the C condition, is consistent with expectations from the frequency theory. Planned comparisons with the C condition, via Dunnett's test, however, revealed that only the difference between the means of the AW-NI and the C condition exceeded the critical value ($p < .05$) of 7.60.

*Errors to the base of opportunities.*—Seven *S*s of the total of 60 failed to reach the 2-perfect trial criterion within 15 trials, 2 in each of the AR-NI and AW-I groups and 1 in each of the remaining three groups. The 2 × 2 analysis of variance on the proportion of errors to the base of opportunities again resulted in a significant effect for type of IAR lists, $F(1, 56) = 9.58$, $p < .005$, and for instructions, $F(1, 56) = 4.71$, $p < .05$. The means and standard deviations for the cells were .165, .047 (AR) and .202, .043 (AW); .171, .046 (I) and .196, .050 (NI). The interaction yielded $F < 1$. The simple analysis of variance including the C group was also significant, $F(4, 55) = 4.03$, $p < .01$. The mean proportions and standard deviations were .153, .045 (AR-I), .158, .033 (C), .177, .050 (AR-NI), .188, .042 (AW-I), and .216, .045 (AW-NI). Comparisons with the C group via Dunnett's test revealed that again only the difference between AW-NI and the C group attained significance, with the critical value being .038.

Experiment II therefore succeeded in illustrating the overall superiority of AR lists to AW lists and in obtaining the predicted inferiority of the AW-NI condition to the C condition. It, however, again failed to find the AR condition significantly superior to the C condition, although the obtained differences were generally in the expected direction. Instructions facilitated the learning of both types of IAR lists but not significantly so relative to the C condition.

## EXPERIMENT III

Experiment II indicated that even with instructions to *S*s regarding the composition of the AR and AW lists, the IAR mechanism postulated by frequency theory was apparently not evoked with sufficient strength to differentiate the AR-I and AW-I groups from the C condition. This result for interitem list manipulations is consistent with the null results for instructions reported by Eberlein and Raskin (1968) for single-list intrapair and interpair IAR-list manipulations, but casts no enlightment upon the problem of replicating the results of Ekstrand et al. (1966) or of explicating the mixed results of Kanak et al. (1970). Of particular concern was the fact that in none of the four experiments thus far in our own laboratory had we replicated the Ekstrand et al. finding of significant superiority of the AR condition to the C condition, although consistent with the frequency theory, we had found the AW condition inferior to the C condition in two of the five comparisons on errors on the first four trials and the AR condition numerically superior on most occasions to the C condition.

Although Exp. I of the present series suggested that certain methodological factors (i.e., word materials, method of presentation, and conduct of the initial trial) were unrelated to performance in the three list conditions, there still remains at least one potentially important methodological difference between the series of studies in our own laboratory and the Ekstrand et al. (1966) study. All of our own studies had employed the visual mode of presentation, while Ekstrand et al. used the aural mode. Although even instructions (Exp. II) with visual presentation apparently failed to evoke the IAR processes predicted by frequency theory, it is conceivable that aural presentation might more readily elicit IARs than visual presentation, particularly since as yet unpublished data in our own laboratory comparing the two Palermo and Jenkins (1964, visual presentation; 1966, aural presentation) normative studies indicated that for the only grade level common to both norms, fourth-grade children do show a higher mean frequency of primary associates when the materials are presented aurally. Although empirical data are needed to determine if this result is also true for adults, the possibility seemed plausible enough to warrant studying mode of presentation in Exp. III. In addition, though apparently no studies have as yet demonstrated an interaction of mode of presentation with other variables, Schulz and Hopkins (1968) have reported superiority for the aural mode.

Since the possibility also exists that any resultant tendency of the two modes to elicit IARs with differential strength might interact with method of presentation (i.e., ANT versus ST), the latter was again included as a factor. Finally, both the original list materials of Ekstrand et al. (1966) and our own derived lists of Exp. I and II were retained because of their potential for interaction with the mode of presentation, particularly in the AR and AW conditions. With the 3 list conditions, Exp. III involved a 3(List Conditions) × 2(Method of Presentation) × 2(Mode of Presentation) × 2(List Materials) factorial design. To approximate as closely as possible the experimental conditions of Ekstrand et al. in the aural presentation-ST procedure, the initial trial was preceded by a study trial for both the ANT and ST conditions.

### *Method*

*Subjects.*—A total of 120 introductory psychology students served as *S*s and none had prior experience in a VD learning experiment. The *S*s were randomly assigned to the 24 ($N = 5$) experimental conditions within the restriction of maintaining equal *N*s across cells before $N + 1$ *S*s were assigned to any cell.

*Procedure.*—Standard instructions and examples consistent with the requirements of the ANT and ST procedure were given to each *S*. The lists were presented via a Lafayette memory drum at a 2:2-sec. rate for *S*s receiving the visual mode of presentation. Under the ANT method, the anticipation and feedback intervals involved the same procedural characteristics as in the previous experiments.

Under the aural mode of presentation, the materials were presented via a tape recorder for both methods of presentation. During the anticipation interval for the ANT condition and the test trials of the ST conditions, the items composing a given pair were presented once and *S* responded with his choice within the 2-sec. interval. In the feedback intervals of the ANT condition and the study trials of the ST condition, the pair was pronounced from the tape and the R item was repeated and labeled "correct" (e.g., "Heavy-low, Heavy is correct") during the 2-sec. interval. As noted earlier, *S*s learning via the ANT procedure were given an initial study trial to equate their experience with the materials to that of the *S*s of the ST condition prior to the first performance trial.

The intertrial interval for all conditions of presentation was 4 sec. Controls for serial position cues again involved four unsystematic serial orders, but the orders were independently randomized for study versus test trials in the ST condition. Under all four combinations of mode and method of presentation, spatial or temporal position of the R item was randomized within the restrictions that across the four serial orders a given R item occurred twice in each spatial or temporal position. The randomizations of spatial positions were identical for all four mode-method combinations.

### *Results*

*Errors on Trials 1–4.*—Seventeen *S*s achieved criterion in fewer than four trials, with 6, 5, and 6 *S*s in the C, AR, and AW conditions, respectively. The analysis of variance on the errors over Trials 1–4 indicated significant effects for list materials and method of presentation, $F\ (1, 96) = 5.11$, $p < .05$, and $23.67$, $p < .005$. Both the Ekstrand et al. (1966) list ($\bar{X} = 8.85$) and the ST method ($\bar{X} = 7.20$) resulted in fewer errors than the derived list ($\bar{X} = 11.65$) and the ANT method ($\bar{X} = 13.37$). The main effect for list conditions was nonsignificant ($F < 1$) with means and standard deviations of 9.80, 8.59 (AR), 9.98, 6.73 (C), and 11.08, 7.99 (AW). No other main effects or interactions were significant, although the Mode of Presentation × Method of Presentation × List Conditions interaction approached significance, $F\ (2, 96) = 2.87$, $p < .10$.

*Errors to the base of opportunities.*—Only five *S*s failed to attain criterion in 15 trials, but four of these were in the AR-ANT-Visual Mode condition. The analysis of variance of proportion of errors to the base of opportunities again produced significant main effects for method of presentation, $F(1, 96) = 22.16$, $p < .001$, and for list materials, $F\ (1, 96) = 5.06$, $p < .05$. List conditions, however, again resulted in $F < 1$ with means and standard deviations of .077, .045 (C), .089, .059 (AR), and .089, .061 (AW). The List Materials × Method of Presentation interaction was significant $F\ (1, 96) = 4.26$, $p < .05$, with the two lists producing no difference under the ST method but the derived list was somewhat more difficult under the ANT method. More importantly, the three-way List Conditions × Method of Presentation × Mode of Presentation interaction, which produced only a statistical trend in the previous analysis, was significant, $F\ (2, 96) = 3.71$, $p < .05$. No other main effects or interactions approached significance.

Table 1 presents the mean proportions for the cells involved in the significant second-order interaction. Tests of the simple-main effects for list conditions at levels of method and mode of presentation resulted in only one significant comparison which reflected significant differences among the list conditions under the visual presentation—ANT method procedure, $F\ (2, 96) = 3.22$, $p < .05$. Table 1 shows that the AR condition resulted in the poorest performance under these conditions. The simple-simple main effects for each list condition across methods and modes of presentation resulted in a significant difference for only the AR condition, $F\ (1, 96) = 7.04$, $p < .01$. The AR list performance was poorest under ANT-visual presentation, followed by ANT-aural, with only minor differences between modes of presentation under the ST method. These results tentatively suggest that method and mode-of-presentation variables may at least partially account for the discrepancies between the findings of Kanak et al. (1970) and the results of Ekstrand et al. (1966). However, this conclusion must be embraced with caution since, as will be seen, Exp. IV of the present series, which used the ANT method and visual presentation, found the AR condition significantly superior to each of the other list conditions and Exp. I and II found the AR condition, with visual presentation, ranking superior to the AW and C conditions. The discrepant result in the present experiment is likely a result of differential difficulty of list materials or of chance factors in the assignment of *S*s to conditions, even though the four-way interaction was clearly nonsignificant, since the mean ($N = 5$) errors to the base of opportunities for the AR condition was .066 under the Ekstrand et al. list versus .098 under the derived list.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR THE LIST CONDITIONS × METHOD OF PRESENTATION × MODE OF PRESENTATION INTERACTION (EXPERIMENT III) FOR TOTAL ERRORS TO THE BASE OF OPPORTUNITIES

| List cond. | Method and mode of presentation | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Anticipation | | | | Study-Test | | | |
| | Visual | | Aural | | Visual | | Aural | |
| | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* |
| C | .089 | .041 | .108 | .028 | .067 | .046 | .046 | .046 |
| AW | .104 | .046 | .098 | .101 | .069 | .036 | .085 | .040 |
| AR | .141 | .070 | .084 | .041 | .040 | .026 | .062 | .022 |

Note.—C = control; AW = associatively related wrong items and unrelated right items; AR = a list of associatively related right items and unrelated wrong items.

## EXPERIMENT IV

The results of Exp. III generally failed to support the frequency theory even though the rank of the list conditions in the analysis of errors on Trials 1–4 was consistent with the theory. Experiments I and III have therefore indicated that although certain of the methodological variables manipulated (e.g., method of presentation, list materials, guessing vs. no guessing) affect the proficiency of VD learning, these variables do not differentially affect performance in the three list conditions, if one assumes that the second-order interaction of Exp. III is primarily an artifact.

In Exp. IV, an attempt was made to further maximize the conditions for the operation of IARs as predicted from frequency theory by again manipulating instructions, but this time with a concurrent manipulation of rate of presentation. Ekstrand et al. (1966) used a 3-sec. rate of presentation, while all the studies in our laboratory had employed a 2-sec. rate. Several investigators (e.g., Ingison & Ekstrand, 1970; Kanak, 1968; Wike & Wike, 1970) have reported that VD learning proceeds more rapidly with longer durations of exposure time. Since Exp. II indicated that instructions facilitated AR and AW list performance, but not significantly so relative to the C condition, it is conceivable that a slower rate of presentation combined with instructions concerning the associated items might produce significant differences between the IAR lists and the C condition.

The present study used 2:2- and 4:4-sec rates, along with the presence or absence of instructions regarding the list relationships for all three list conditions. The AR and AW conditions were given instructions comparable to those employed for the instruction conditions of Exp. II. In the C condition, "instructed" *S*s were told that there were no items in the list which were associatively related and that the words were chosen for this reason. The C-I condition was included since little is known regarding the effects of any type of instructions for a control group beyond those standardly used, which typically give no information regarding attributes of list items but, rather, are limited to procedural and response characteristics of the task. In addition, Exp. IV maintained equal numbers of *S*s within each sex in order to examine the possibility of sex-related differences in the effects of the dependent variables. Two studies (Fulkerson & Johnson, 1971; Schulz & Hopkins, 1968) have reported that females learned a VD task more easily than males, but little is known of the possible interaction of task variables with the sex of *S*s. The experimental design was thus a 3 (List Conditions) × 2 (Rate of Presentation) × 2 (Instructions) × 2 (Sex) factorial.

## *Method*

*Subjects.*—A total of 144 introductory psychology students participated in the experiment. The 12 treatment conditions were randomly assigned to *S*s upon their appearance at the laboratory within the restriction of maintaining an equal sex distribution within each condition. None of the *S*s had participated in previous verbal learning experiments.

*List materials and instructions.*—The list materials were the pairs from the original lists employed by Ekstrand et al. (1966) modified in a minor way as reported in Exp. I. As in Exp. II, *S*s of the AR-I and AW-I conditions were given explicit information regarding the associative relationships and the particular item function served by the associates. The *S*s of the C-I condition were given instructions calling their attention to the absence of associative relationships among the words. Instructions for the AR-NI, AW-NI, and C-NI conditions were similar to the uninstructed conditions for Exp. II.

*Procedure.*—All *S*s were tested individually. The lists were presented by the anticipation method, using an initial guessing trial. With the exception of the variation of a 2:2-versus a 4:4-sec. rate of presentation, all details of the procedure were identical to those of the previous experiments. The criterion was again two successive errorless trials or 15 trials, whichever occurred first.

## *Results*

*Errors on Trials 2–5.*—The list conditions differed significantly on errors over the first four trials, $F\ (2,\ 120) = 14.51$, $p < .001$. The means and standard deviations, respectively, were 10.96, 8.35 (AR), 18.08, 10.96 (AW), and 20.81, 9.31 (C). However, although instructions failed to produce a main effect ($F < 1$), the List Conditions × Instructions interaction was significant, $F\ (2,\ 120) = 5.70$, $p < .025$. The means involved in the interaction are presented in Table 2. The effects for instructions upon the AR and C lists were both significant, $F$s (1, 120) = 5.03 and 5.10, respectively, $p$s $< .05$. Instructions significantly facilitated AR list performance, but the C condition was apparently adversely affected by the information that the list contained no associatively related words. The latter result was unexpected and does not have a readily apparent explanation. The presence or absence of instructions did not significantly affect performance in the AW condition, although in agreement with Exp. II, the AW-I condition was numerically superior to AW-NI.

The differences between the list conditions were highly significant under instructions, $F\ (2,\ 120) = 17.63$, $p < .001$, but the differences under no instructions only approached significance, $F\ (2,\ 120) = 2.58$, $p < .10$. Newman-Keuls test indicated that, in accordance with frequency theory, the AR-I condition was superior to each of the other two instructed conditions ($p < .01$), but contrary to the theory, the AW-I condition was also significantly superior to C-I ($p < .01$). Since the performance of the C-I condition was unexpectedly poor and little is known of the effects of such instructions in a control condition, comparisons between list conditions under the No-Instruction condition were also undertaken despite the fact that these differences failed to reach the usually accepted level of significance. A Newman-Keuls test indicated that only the AR-NI versus AW-NI comparison was significant ($p < .05$). Identical conclusions were reached with respect to the nonsignificant comparison of the two experimental conditions with the C condition, employing the more powerful Dunnett's test. The rank of the means relative to the C condition is in accord, however, with the frequency theory.

The 2:2-sec ($\bar{X} = 18.79$) rate of presentation resulted in more errors than the 4:4-sec. rate ($\bar{X} = 14.47$), $F\ (1,\ 120) = 7.83$, but the interactions of rate of

TABLE 2

Means and Standard Deviations for the List Conditions × Instruction Interaction (Experiment IV) for Errors on Trials 2–5

| IAR instructions | List Cond. | | | | | |
|---|---|---|---|---|---|---|
| | C | | AW | | AR | |
| | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* | $\bar{X}$ | *SD* |
| Instructions | 23.83 | 8.81 | 16.29 | 11.09 | 7.96 | 7.24 |
| No instructions | 17.79 | 8.98 | 19.96 | 10.76 | 13.96 | 8.45 |

Note.—IAR = Implicit associative response; C = control; AW = a list of associatively related wrong items and unrelated right items; AR = a list of associatively related right items and unrelated wrong items.

presentation with list conditions and instructions, as well as the second-order interaction, were all nonsignificant (*F*s $< 1$). Males ($\bar{X} = 18.46$) made more errors than females ($\bar{X} = 14.80$), $F\ (1,\ 120) = 5.60$, $p < .025$, but sex produced no significant interactions with other factors (all *F*s $< 1$).

*Total errors to the base of opportunities.*—A total of 14 *S*s of 144 failed to reach 2 perfect trials within 15 trials. List conditions also differed significantly in the error rate measure, $F\ (2,\ 120) = 9.04$, $p < .01$; but the interaction with instructions, in contrast to the above analysis, failed to attain significance, $F\ (2,\ 120) = 2.99$, $p < .10$. The main effect for instructions was also nonsignificant $F\ (1, 120) = 1.82$. The mean proportions and standard deviations for the list conditions were .152, .049 (AR), .189, .058 (AW), and .191, .049 (C). Newman-Keuls test indicated that the AR condition was significantly (*p*s $< .01$) superior to each of the other conditions which did not differ from each other. The more powerful Dunnett's test confirmed that the AW versus C comparison was nonsignificant. The List Conditions × Rate of Presentation × Sex interactions both yielded $F < 1$.

## Discussion

In general, the evidence in support of the frequency theory for interitem manipulations across the four experiments was inconsistent. On the measure of errors over the early trials, a measure on which Ekstrand et al. (1966) found the AR condition superior to the C condition as predicted, only in Exp. IV was this result replicated, but the superiority was present under both noninstructed and instructed conditions. The AR condition was found to be significantly superior to the AW condition in Exp. II and Exp. IV, however, consistent with the theory. In Exp. I and III, the AR condition was numerically superior to the other conditions but not significantly so. The AW condition, which Ekstrand et al. found did not differ significantly from the C condition, was found to be significantly inferior to the C condition, as the theory predicts, only in the noninstructed condition of Exp. II and was significantly inferior to the AR condition only in Exp. II and IV. The rank of the AW condition, relative to the C and AR conditions, was consistent with the theory only in Exp. II and III, although the AW condition was also numerically inferior in the noninstructed conditions of Exp. IV. The measure of errors to the base of opportunities generally yielded the same conclusions as the measure of errors on the early trials, with the exception of some switches in the rank of the groups in Exp. II, III, and IV.

While the above summary of results for the four experiments suggests some support for the frequency theory, the inconsistencies are perhaps as outstanding. These inconsistencies, combined with the two experiments of Kanak et al. (1970) and the null results for the AW condition in the experiment of Ekstrand et al. (1966), suggest that the predicted function of IARs in interitem manipulations may be called into some question. In addition, the predicted role of IARs in *intrapair* manipulations has already been questioned since others (e.g., Barch et al., 1967; Fulkerson & Kausler, 1969) have found, contrary to the theory, that intrapair manipulations yield superiority to, or no difference from a C condition. Intrapair manipulations have been interpreted by the above authors as allowing the activation of higher order rules or strategies based on IARs which mitigate the reduction of the differential frequency cue stemming from IAR activation per se. The evidence presented by the above authors does not question the activation of IARs, but rather suggests that IARs allow the activation of processes compensatory to the reduction of frequency cues presumed by the frequency theory to be occasioned by IARs. In the interpair condition, on the other hand, such a compensatory mechanism may be inoperable and the available literature (e.g., Eberlein & Raskin, 1968; Fulkerson & Kausler, 1969) indicates that the predicted inferiority of interpair conditions to the C condition does obtain. The compensatory process involving IARs in the intrapair manipulation and the actual reduction of the frequency cue in the interpair manipulation may both be stimulated by the cue provided by the item function contrast (i.e., associated items serve opposing W and R functions) with resultant activation of facilitation and interference mechanisms, respectively.

In contrast, the interitem manipulation, involving associative relationships only within W items or within R items rather than between them, may result in a much weaker activation of IARs sensitive to only a limited range of variables such as instructions. Instructions facilitated the learning of both types of IAR lists in Exp. II, but not significantly so relative to the control condition. In Exp. IV, instructions significantly facilitated performance of the AR condition, relative to no instructions, and improved the performance of the AW condition to a nonsignificant degree relative to the noninstructed AW condition, as in Exp. II. The latter results suggest that instructions in the AW condition may activate a higher order strategy mitigating the reduction of the differential frequency cue. One such strategy could be the avoidance of associatively related items and the consequent selection of the alternative. It is interesting to note that Eberlein and Raskin (1968) found no effect for instructions in interpair and intrapair manipulations, which suggests that item function contrast may have served as a sufficient cue to activate attention to IARs.

The evidence presented in support of the frequency theory, although inconsistent, suggests that IARs may indeed operate in interitem manipulations, but with less strength and more variability than expected. As mentioned earlier, one source for such a deficit in strength may be the fact that the related items serve the same function and hence draw less attention than related items in intrapair or interpair manipulations. Several other potential sources for the greater weakness of IAR mechanisms in interitem manipulations may be identified. For example, the overt pronunciation of R items may provide a sufficiently distinctive cue in the AW condition to overcome the reduction of the frequency cue produced by IARs to W items. Similarly, in the AR condition, the overt pronunciation response may be a sufficient condition to facilitate learning without the evocation of IARs. In addition, the weaker degree of W item learning qua response, relative to R item learning (Kausler & Sardello, 1967), implies that IAR effects in the AW condition would likely be weaker than in the AR condition when both are evaluated against the C condition. The findings of Ekstrand et al. (1966) and of the present experiments are consistent with this notion. Furthermore, the absolute effect of IARs may be expected to be relatively subtle, particularly in the absence of a contrast of related item function, since the normative values (Palermo & Jenkins, 1964) for frequency of the primary associate to a stimulus item rarely exceed 50%–60% of the *S*s. That is, for a substantial porportion of *S*s learning an IAR list the "primary" associate as given by the norms may indeed be a secondary associate. Indeed, Kausler and Olson (1969) have shown strong support for the frequency theory in both intrapair and interpair manipulations when the R or W items are related through the acoustic similarity inherent in homonyms, materials which control across *S*s the degree of similarity being manipulated. The composite effects of these factors may account for the seemingly weak effects of IARs in interitem manipulations.

Finally, with regard to the effects of methodological variables on the learning of a VD list of associatively unrelated items (i.e., the C condition), the results of the present studies agree with others cited earlier in demonstrating more efficient learning (*a*) under the study-test method as opposed to anticipation, (*b*) with no guessing on the initial trial, (*c*) and under slower rates of presentation, but with females learning faster than males under a 2:2-sec. rate. In view of the lack of significant interactions involving these factors, these conclusions may apparently be generalized to the AR and AW interitem IAR manipulations.

## REFERENCES

Barch, A. M., Lippman, M. & Whalen, S. Intrapair relations in verbal discrimination learning: A paradoxical finding. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1967.

Battig, W. F., & Switalski, R. W. Comparison of anticipation and recall procedures in verbal discrimination learning. *Psychonomic Science*, 1966, 6, 65–66.

Bousfield, W. A., Whitmarsh, G. A., & Danick, J. J. Partial response identities in verbal generalization. *Psychological Reports*, 1958, 4, 703–713.

Dominowski, R. L. First-trial guessing and VD learning. *Psychonomic Science*, 1966, 5, 231–232.

EBERLEIN, E., & RASKIN, D. C. Intrapair and interpair associations in verbal discrimination learning. *Psychonomic Science*, 1968, **11**, 145–146.

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

FULKERSON, F. E., & JOHNSON, J. E. Methodological variables in verbal discrimination learning. *Psychonomic Science*, 1971, **22**, 68–69.

FULKERSON, F. E., & KAUSLER, D. H. Supplementary report: Effects of intra-pair and inter-pair bidirectional associates on verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 307–310.

INGISON, L. J., & EKSTRAND, B. R. Effects of study time, method of presentation, word frequency, and word abstractness on verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **85**, 249–254.

KANAK, N. J. The effects of rate of exposure upon simultaneous intentional and incidental verbal-discrimination learning. *Psychonomic Science*, 1968, **12**, 141–142.

KANAK, N. J., COLE, L. E., & THORNTON, J. W. Inter-item manipulations of implicit associative responses in verbal discrimination learning. *Psychonomic Science*, 1970, **18**, 359–361.

KAUSLER, D. H., & OLSON, R. D. Homonyms as items in verbal discrimination learning and transfer. *Journal of Experimental Psychology*, 1969, **82**, 136–142.

KAUSLER, D. H., & SARDELLO, R. J. Item recall in verbal discrimination learning as related to pronunciation and degree of practice. *Psychonomic Science*, 1967, **7**, 285–286.

PALERMO, D. S., & JENKINS, J. J. *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press, 1964.

PALERMO, D. S., & JENKINS, J. J. *Oral word association norms for children in grades one through four.* (Research Bulletin No. 60) University Park, Pennsylvania State University, 1966.

SCHULZ, R. W., & HOPKINS, R. H. Presentation mode and meaningfulness as variables in several verbal-learning tasks. *Journal of Verbal Learning and Verbal Behavior*, 1968, **1**, 1–13.

WIKE, S. S., & WIKE, E. L. Presentation rates and verbal discrimination learning. *Psychonomic Science*, 1970, **19**, 253–254.

(Received September 7, 1971)

Journal of Experimental Psychology
1972, Vol. 93, No. 2, 320-326

# REPETITION EFFECT AS A FUNCTION OF EVENT UNCERTAINTY, RESPONSE–STIMULUS INTERVAL, AND RANK ORDER OF THE EVENT[1]

CARLO UMILTÁ,[2] CHARLES SNYDER,[3] AND MARTHA SNYDER

*University of Oregon*

In reaction time (RT) studies it has been found that RTs for repeated events may be faster (positive recency effect or repetition effect) or slower (negative recency effect) than RTs for nonrepeated events. The present experiment was designed to investigate how the positive and negative recency effects vary with: (*a*) the event uncertainty; (*b*) the delay between a response and the next stimulus (RSI); (*c*) the rank order of the event in a sequence of equivalent events. Our results show that: (*a*) under the conditions of this study there is either a positive recency effect or no effect at all; (*b*) the effect is larger when uncertainty is greater; (*c*) the effect diminishes as RSI increases; (*d*) the effect is larger at the beginning than at the end of a sequence of equivalent events when event uncertainty is low, while the opposite is true when event uncertainty is high. The importance of these and other findings with respect to memory and expectancy hypotheses for the repetition effect is discussed.

It is well known that in discrete (Hyman, 1953) and in serial (Bertelson, 1961) reaction time (RT) tasks, RTs to repeated signals are usually shorter than RTs to signals which are new (i.e., different from the immediately preceding one). Bertelson named this phenomenon the repetition effect and suggested two principle explanatory schemes, the "memory hypothesis" and the "expectancy hypothesis."

According to one version of the memory hypothesis, the repetition effect might be due to some peripheral motor facilitation lasting for a short period of time and causing a second identical response to be made more quickly. Alternatively, the effect might result from residual activity in S–R bonds which would eliminate the need for a new memory search when a second identical stimulus occurred. In this formulation, the repetition effect would be a particular instance of a short-term memory effect. As such, the memory hypothesis gains support from those studies which have shown that the repetition effect, like short-term memory, is reduced as the interval between a response and the next stimulus (RSI) is increased, and particularly for incompatible S–R arrangements (Bertelson & Renkin, 1966).

These results, however, have not been confirmed by Schvaneveldt and Chase (1969) or by Keele (1969) for compatible or for incompatible S–R arrangements. Furthermore, an inverse relationship between repetition effect and RSI, even if real, would not be inconsistent with an expectancy hypothesis. This notion maintains that *S* always prepares for certain events, namely for repetitions, and reacts faster when his expectation is confirmed than when it is disconfirmed. The saving is in memory retrieval time when an event is correctly anticipated (Keele, 1969; Schvaneveldt & Chase, 1969). It is possible to combine the memory and expectancy hypotheses by suggesting that the memory of a previous event acts like an irrelevant stimulus, in the sense of Greenwald (1970), which causes *S* to expect a repetition. In this formulation, "event" would mean either a particular stimulus, a particular response, or the combination; in short, whatever might be causing the repetition effect.

---

[1] This research was supported by the Advanced Research Projects Agency for the Department of Defense and was monitored by the United States Air Force Office of Scientific Research under Contract F 44620-67-C-0099. We wish to thank S.W. Keele and R. Hyman for helpful suggestions during the preparation of the manuscript.

[2] Now at the Institute of Psychology of the University of Bologna, Italy.

[3] Requests for reprints should be sent to Charles Snyder, University of Oregon, Department of Psychology, Eugene, Oregon 97403.

It has also been shown (Bernstein & Reese, 1965; Keele, 1969) that when *S*'s anticipation of a stimulus is correct, RTs are faster than when *S*'s anticipation is incorrect. According to this version of the expectancy hypothesis, *S*s have a tendency to anticipate the recurrence of an immediately preceding event, making event repetitions confirmatory and, therefore, fast (Keele, 1969; Schvaneveldt & Chase, 1969). Furthermore, Bernstein and Reese (1965) found the difference between RTs following correct and incorrect anticipations to be a direct function of event uncertainty. These findings could explain Hyman's (1953) result that the more alternative events, the greater the repetition effect. They would also shed light on the demonstration by Hyman and others (e.g., Schvaneveldt & Chase, 1969; Williams, 1966) of a negative recency effect. These studies, using only two equiprobable events with relatively long RSI and compatible S–R arrangements, produced a tendency to react faster to an event different from the immediately preceding one (the so-called gambler's fallacy is thought to be a special instance of this effect). Thus, the expectancy hypothesis suggests that a negative recency effect depends upon high event probability, long RSI, and compatible S–R arrangement. Some results in support of these suggestions have already been reported (Umiltá & Trombini, 1968).

The present experiment was designed to investigate both negative and positive recency effects and how they vary with (*a*) event probability or uncertainty, (*b*) length of RSI, and (*c*) rank order of an event in a sequence of equivalent events.

## METHOD

*Subjects.*—Sixty *S*s (12 female) with normal or corrected-to-normal vision were obtained through the University of Oregon Student Employment Service and paid $1.50/hr. Each *S* was run individually at the same hour for 2 consecutive days, in sessions lasting about 30, 60, or 90 min. according to length of RSI.

*Apparatus.*—Stimulus presentation, timing, and response recording were under the control of a PDP-9 computer. Stimuli were presented on an in-line readout display (Industrial Electronic Engineers Model 10011-44-B) situated in a separate experimental room. On each trial, one of the four digits (1, 2, 3, 4) used as stimuli was presented in the same spatial location. Directly beneath the display was a four-key response board on which *S*s placed the middle and index fingers of each hand. The plastic keys operated microswitches requiring a force of about 400 gm. Designating the keys A–D from left to right, the stimulus–response mapping was 1-A, 2-B, 3-C, and 4-D. Reaction time was measured from the onset of a digit to the first key press, although the correction procedure used required the digit to remain lighted until the correct response was made.

*Design.*—Two stimulus order conditions were used, each condition consisting of four alternatives with equal frequencies of occurrence. For both conditions, the average information per trial was .97 bits. In Cond. $C_1$ (high repetition), the probability that a stimulus would repeat itself on the next trial was .82, while the probability of occurrence of each of the three possible alternations was .06. In Cond. $C_2$ (high alternation), there were, after every stimulus, one high-probability alternation (.82), two low-probability alternations (.06), and one low-probability repetition (.06). Transitional probability assignments in Cond. $C_2$ yielded six high-probability event sequences: 1-2-3-4, 1-2-4-3, 1-3-4-2, 1-3-2-4, 1-4-2-3, and 1-2-1-2, or 3-4-3-4. Condition $C_2$ *S*s were randomly assigned to one of these sequences.

Three RSIs of 250, 2,000, and 3,750 msec. were employed in each of the two stimulus order conditions (i.e., high-probability repetitions and high-probability alternations). Ten *S*s were randomly assigned to each of the six stimulus order by RSI conditions.

*Procedure.*—Twenty blocks of 50 trials were presented in each experimental session. Following each block, *S*s were given a 20-sec. rest period during which they could read from an oscilloscope their mean RT and number of errors for the preceding block. Between Blocks 10 and 11, *S*s were allowed a 5-min. rest interval.

The first 10 blocks of Day 1 were used to familiarize *S*s with equipment and task procedures, and those data were not included in the RT and error analyses below. Before testing on Day 1 commenced, *S*s were instructed to press the key, corresponding to the digit shown on the display, as rapidly as possible, and to try not to make more than two errors during any block of 50 trials. The statistical structure of the stimulus order was also explained, and all *S*s indicated that they understood the sequential probabilities before the session was begun.

## RESULTS

The treatment of RT data involved separate analyses of variance for high-probability stimuli (i.e., repetitions in Cond. $C_1$ and high-probability alternations in Cond. $C_2$) and for low-probability stimuli (i.e., alternations in Cond. $C_1$ and low-probability repetitions in Cond. $C_2$). In addition to these categories of events, reference will sometimes be made to "equivalent" events as those having the same objective probability of occurrence and being of the same response type, repetitions or alternations. The following is an exposition of the results of these analyses considered by source of variability.

There was a significant main effect of RSI, a between-*S*s factor, both for high-probability, $F\ (2, 54) = 7.06, p < .005$, and for low-probability events, $F\ (2, 54) = 11.06, p < .001$. Reaction times increased with length of RSI from 334 to 344 to 375 msec. for high probability, and from 509 to 508 to 565 msec. for low probability.

The main effect of the between-*S*s factor of response type, repetition versus alternation, was also significant for both high-probability, $F\ (1, 54) = 9.76, p < .005$, and low-probability events, $F\ (1, 54) = 30.04, p < .001$. In both cases there was a positive recency effect, or repetition effect, which was greater for low probability (63 msec.) than for high probability (28 msec.).

The within-*S*s main effect of sequential order in a sequence of equivalent events was also significant for high probability, $F\ (1, 54) = 42.70$, and for low probability, $F\ (1, 54) = 37.18$, both *p*s $< .001$. In the case of high probability, this comparison is between the first event and events following the sixth event, in a sequence of equivalent events. For low probability, the comparison is between the first and the second in a sequence of equivalent events, since low-probability sequences were seldom longer than two events. Reaction times are faster to later events in a sequence for both high-probability (52 msec.) and low-probability (27 msec.) events.

The interaction between RSI and response type, repetition versus alternation, also was significant for high probability, $F\ (2, 54) = 3.68$, and for low probability, $F\ (2, 54) = 3.93$, both *p*s $< .05$. The repetition effect diminishes from 61 msec. at the shortest RSI to 25 msec. at the middle RSI and to $-2$ msec. (i.e., there is a very small negative recency effect) at the longest RSI, in the case of high-probability events. The corresponding data for low probability are 98, 65, and 25 msec., respectively, for the shortest, middle, and longest RSI.

The RSI × Sequential Order interaction was significant only for low probability, $F\ (2, 54) = 4.85, p < .025$. In a sequence of equivalent low-probability events, the advantage of the second event over the first diminishes with lengthening RSI from 41 to 32 to 8 msec. The corresponding data for high probability are 61, 51, and 46 msec., $F\ (2, 54) = 2.38, p > .1$.

The interaction between response type, repetition versus alternation, and sequential order was significant for high probability, $F\ (1, 54) = 23.48$, and for low probability, $F\ (1, 54) = 40.00$, both *p*s $< .001$. The change that the repetition effect undergoes in passing from the beginning to the end of a sequence of equivalent events, however, is in opposite directions for the two classes of event probability. For high probability, the repetition effect is greater at the beginning than it is at the end of a sequence (43 and 15 msec.). For low probability, the repetition effect is greater at the end of a sequence (35 and 91 msec.). For high probability, RTs to repetitions decrease less than RTs to alternations, from the beginning to the end of a sequence (39 msec. vs. 67 msec.). For low probability, RTs to repetitions decrease more than RTs to alternations, from the beginning to the end of a sequence (56 msec. vs. 1 msec.).

The RSI × Response Type × Sequential Order interaction was also significant for high probability, $F\ (2, 54) = 35.92, p < .001$, and for low probability, $F\ (2, 54) = 6.89, p < .005$. The effect indicates a change over increasing RSI in the relationship between repetition effect and sequential order. The data are presented in Table 1.

So far this analysis has failed to consider some data, namely low-probability alternations in Cond. $C_2$. A within-*S*s comparison between low-probability repetitions and low-probability alternations of Cond. $C_2$,

TABLE 1

Repetition Effect (in msec.) as a Function of Event Probability, Position in a Sequence of Equivalent Events, and the Interval between a Response and the Next Stimulus (RSI)

| Repetition effect | High probability (.82) | | | Low probability (.06) | | |
|---|---|---|---|---|---|---|
| | First in sequence | After sixth in sequence | Change over sequence | First in sequence | Second in sequence | Change over sequence |
| RSI | | | | | | |
| 250 | 107 | 14 | −93 | 58 | 138 | 80 |
| 2000 | 34 | 22 | −12 | 27 | 103 | 76 |
| 3750 | −12 | 10 | 22 | 20 | 30 | 10 |

however, confirmed those results which have been already shown and did not supply new information.

In order to analyze error data, similar analyses have been performed, only disregarding the factor of sequential order and considering arc-sine transformed percentages instead of raw data.

1. The main effect of RSI was significant only in the case of low probability, $F$ (2, 54) = 6.50, $p < .005$, owing to a slight tendency of the percentage of errors to increase with shorter RSI. The data from the shortest to the longest RSI are: 22.92%, 16.34%, and 19.70% for low probability and .64%, .32%, and .32% for high probability, $F$ (2, 54) = 1.67, $p > .1$.
2. The main effect of response type, repetition versus alternation, was significant only for high probability, $F$ (1, 54) = 27. 62, $p < .001$, and shows that errors are more likely in the case of alternations (.70% vs. .18%). The corresponding data for low probability are 20.52% and 18.78%, $F$ (1, 54) = .26, $p > .2$.
3. The RSI × Response Type interaction was significant only for high probability, $F$ (2, 54) = 10.12, $p < .001$. The data show that errors are more likely to occur in the case of alternations, but mostly at the shortest RSI. The following percentages were found, from the shortest to the longest RSI: 1.27% vs. .06%, .38% vs. .27%, and .44% vs. .20%. The corresponding data for low probability are: 24.24% vs. 21.60%, 15.52% vs. 17.43%, and 22.08% vs. 17.32%, $F$ (2, 54) = .66, $p > .2$.

## Discussion

These results confirm those of previous research, which has shown that the difference between repeated and nonrepeated events declines with increasing RSIs and that there is an increase in RTs to both repeated and nonrepeated events with increasing RSIs. Furthermore, the results show that the repetition effect diminution with increasing RSIs is mostly due to a comparatively greater increase of RTs to repetitions. Reaction times to repeated events are in fact 303, 330, and 376 msec., from the shortest to the longest RSI, for high probability (an increase of 73 msec.); and 460, 475, and 553 msec. for low probability (an increase of 93 msec.). The corresponding RTs to nonrepeated events, again from the shortest to the longest RSI, are 364, 355, and 374 msec. for high probability (an increase of 10 msec.); and 558, 540, and 578 msec. for low probability (an increase of 20 msec.).

These results clearly support the memory hypothesis for the repetition effect since they suggest slow decay in memory of previously activated S–R bonds. In the case of a nonrepeated event, the activation should always be weaker because a comparatively longer time has always elapsed since an occurrence of that event. As we have already suggested, however, this finding is not inconsistent with the expectancy hypothesis, provided that a previously activated S–R bond be thought of as an irrelevant stimulus which causes *S* to tend to check first for a repeated event. The fact that RTs to both repeated and nonrepeated events increase with increasing RSIs can be accounted for by the well-established principle that RT depends on time uncertainty (Bertelson & Tisseyre, 1968), since, from a subjective point of view, the time uncertainty ought to be greater with a 3,750-msec. RSI

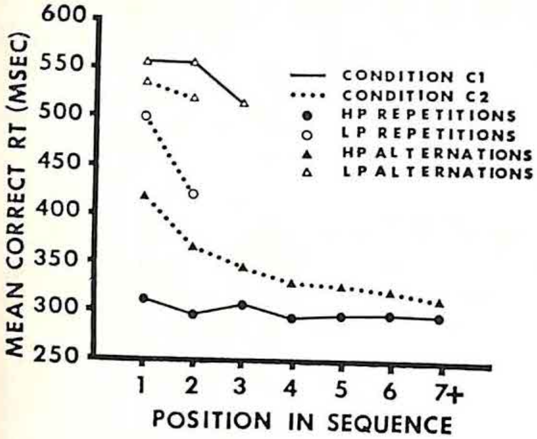


FIG. 1. RSI = 250 msec. (RT as a function of sequential position in series of consecutive equivalent events: high probability (HP) = filled points; low probability (LP) = open points; repetitions = circular points; alternations = triangular points.)

than with a 250-msec. RSI, when no warning signal is given.

The memory hypothesis, however, cannot explain the reason why the repetition effect is larger for low probability than for high probability and why the repetition effect becomes larger in the case of low probability and smaller in the case of high probability at the end of a sequence of equivalent events. On the contrary, as we have already shown, the expectancy hypothesis does suggest a direct relationship between event uncertainty and amount of repetition effect.

Figures 1, 2, and 3 present, for each of the three RSIs, mean RTs (i.e., means taken over mean correct RTs for 10 *S*s) for equivalent events, as a function of sequential order of the events. As these figures show, every function tends to decrease with the lengthening of the sequence of equivalent events, and this seems to suggest another kind of expectancy, namely *S*'s tendency to anticipate the continuation of the sequence in progress.

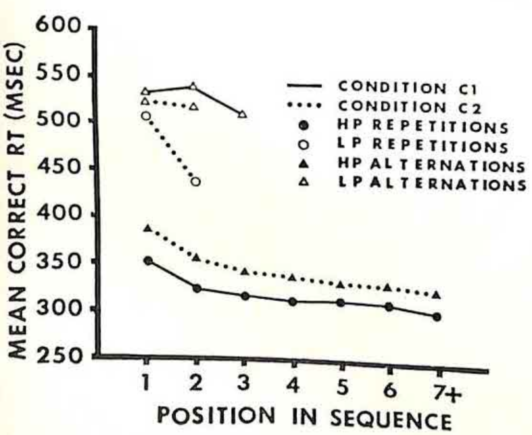


FIG. 2. RSI = 2,000 msec. (RT as a function of sequential position in series of consecutive equivalent events: high probability (HP) = filled points; low probability (LP) = open points; repetitions = circular points; alternations = triangular points.)

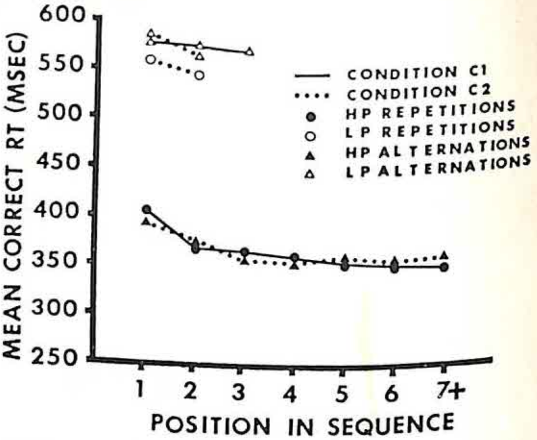


FIG. 3. RSI = 3,750 msec. (RT as a function of sequential position in series of consecutive equivalent events: high probability (HP) = filled points; low probability (LP) = open points; repetitions = circular points; alternations = triangular points.)

With respect to low-probability events, the saving in time brought about by this expectancy is much larger for repetitions than it is for alternations because in the former case *S* is able to prepare for a single event as opposed to two (Cond. $C_2$) or three (Cond. $C_1$) equiprobable events. This is the reason why the repetition effect becomes larger toward the end of the low-probability sequence. Also, the second consecutive alternation in Cond. $C_2$ (where there are two equivalent low alternations) yields a larger saving in time than the second consecutive alternation in Cond. $C_1$ (where there are three equivalent low-probability alternations). At longer RSIs, the tendency to anticipate the continuation of the sequence in progress seems to be overcome, with respect to low-probability events at least, by *S*s' utilization of objective event probability.

Some evidence for this kind of expectancy, which is not visible in Fig. 1, 2, and 3, is the slowing of RTs to high-probability events following a single low-probability repetition. In Cond. $C_2$, for example, the second low-probability repetition is considerably faster than the first, while the second low-probability alternation is not much faster than the first.

This difference diminishes as RSI increases. If fast RTs to the second low-probability repetition are due to *S*s' anticipation of the sequence continuing, then RTs to a high-probability event following a single low-probability repetition should be slowed. At RSI of 250 msec., RTs to a high-probability event are 15 msec. faster following a single low-probability alternation than following a single low-probability repetition. At RSI of 2,000 msec. this difference is 13 msec., and at RSI of 3,750 msec. the difference is −1 msec.

The expectancy hypothesis should thus be modified in such a way as to include *S*s' tendency to expect the repetition of the same event and, in addition, his tendency to expect the repetition of an equivalent event, i.e., the continuation of the sequence in progress.

This hypothesis seems at first sight to have some difficulties with the results for high-probability events. Consider the case when event uncertainty is very low for either alternations or repetitions, i.e., when *E*'s instructions prompt *S* to expect with the same high likelihood both alternations and repetitions. The hypothesis suggests that there ought to be no repetition effect, or a negative recency effect similar to the so-called gambler's fallacy. Furthermore, it suggests that repetitions and alternations ought to profit in the same way by *S*'s tendency to expect the continuation of the sequence in progress, since in this case there is always only one event to prepare for. Instead, we have found that there is a repetition effect, at the shortest RSIs at least, and that alternations profit more than repetitions by coming late in the sequence. In order to explain these findings, it is necessary to consider the fact that the degree of response–response compatibility in our experiment is lower in the case of high-probability alternations than in that of high-probability repetitions. In other words, it may be that the repetition effect found for high probability at the shortest RSIs is due to difficulty in executing the most probable sequence of nonrepeated events, causing *S*s to need a longer time in order to select each response. With the longest RSI, or in a continuing sequence, there is time available for *S*s to prepare a response so that the repetition effect should disappear. In fact, at longer RSIs and at later points in a sequence of equivalent events, RTs to repetitions (compatible) and to alternations (incompatible) are more nearly equal. If the first and second responses in a high-probability alternation sequence are compared at RSI of 250 msec., the difference is found to be 42 msec. (Fig. 1). At RSI of 2,000 msec. this difference is 30 msec. (Fig. 2), and at RSI of 3,750 msec. the difference is 22 msec. (Fig. 3). Comparing the most compatible sequence (1-2-3-4) with the average for all sequences, at 250-msec. RSI the drop is 12 msec., or 28% of the average; at 2,000-msec. RSI the drop is 18 msec., or 59% of the average; and at 3,750-msec. RSI the drop is 16 msec., or 70% of the average. The longer the RSI, then, the smaller the effect of response–response compatibility on the serial position function. Also, the longer the RSI, the flatter the serial position function for high-probability alternations. This is in contrast to the serial position function for high-probability repetitions, which becomes steeper with increasing RSI. It appears likely that *S*'s tendency to anticipate a repetition accounts for some of this interaction since the tendency to anticipate an event repetition, regardless of the objective probabilities, is presumed to have its greatest effect at short RSI.

## CONCLUSIONS

In conclusion, these data seem to be handled well by five explanatory principles, although the fine detail of how these factors operate is not perfectly understood. The slowing of RT with increasing RSI is attributed to *Principle 1*: that greater temporal uncertainty leads to longer *RT*. The advantage of high-probability over low-probability responses is attributed to *Principle 2*: that *S*s' knowledge of objective event probabilities enables them to respond more quickly to more likely events.

The advantage of repetitions over alternations, especially at short RSI, is attributed to *Principle 3*: that *S*s tend to anticipate an event repetition, regardless of objective event probabilities, for some period of time following each response. It seems that *S*s with a long RSI during which to evaluate objective event probabilities are less apt to demonstrate this recency effect.

The shorter RTs to later events in a sequence of equivalent events are attributable to *Principle 4*: *S*s tend to anticipate the continuation of a sequence, and when such anticipations are confirmed, RTs are fast. At longer RSIs, this tendency, with respect to low-probability sequences, is overcome by *S*s' utilization of objective event probabilities.

Finally, the convergence of the serial position functions for high-probability repetitions and alternations, with increasing RSI, is due to *Principle 5*: compatible response sequences require less time to reach their asymptotic RT than incompatible sequences.

## REFERENCES

BERNSTEIN, I. H., & REESE, C. Behavioral hypothesis and choice reaction time. *Psychonomic Science*, 1965, **3**, 259–260.

BERTELSON, P. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 90–102.

BERTELSON, P. & RENKIN, A. Reaction times to new versus repeated signals in a serial task as a function of response-signal time interval. *Acta Psychologica*, 1966, **25**, 132–136.

BERTELSON, P., & TISSEYRE, F. The time-course of preparation with regular and irregular foreperiods. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 297–300.

GREENWALD, A. G. Selective attention as a function of signal rate. *Journal of Experimental Psychology*, 1970, **86**, 48–52.

HYMAN, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, **45**, 188–196.

KEELE, S. W. Repetition effect: A memory-dependent process. *Journal of Experimental Psychology*, 1969, **80**, 243–248.

SCHVANEVELDT, R. W., & CHASE, W. G. Sequential effects in choice reaction time. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

UMILTÁ, C., & TROMBINI, G. La distribuzione delle probabilitá degli stimoli come fattore determinante nei tempi di reazione a scelta multipla. *Rivista di Psicologia*, 1968, 91–98, (Special Issue).

WILLIAMS, J. A. Sequential effects in disjunctive reaction time: Implications for decision models. *Journal of Experimental Psychology*, 1966, **71**, 665–672.

(Received September 20, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 327–332

# EFFECTS OF ATTRIBUTE PROBABILITY IN A MEMORY SEARCH TASK[1]

JOSEPH DUMAS,[2] ELAINE GROSS
*Oakland University*

AND STEPHEN F. CHECKOSKY
*Lake Forest College*

In two experiments, *S*s were asked to respond either YES or NO that a visually displayed two-dimensional pattern had been contained in a previously memorized set of three patterns. Two of the three patterns shared one high-probability attribute (HP). The structure of the stimulus sets was such that the probability of a YES response given the HP attribute was greater than the probability of a NO response. The results from both experiments showed that reaction times to patterns having the HP attribute were faster than patterns not containing the HP attribute for YES responses but slower for NO responses. It was concluded that the comparison stage in a memory search task is characterized by self-terminating feature testing and that the comparison and response selection stages can occur in parallel.

During the past 20 yr. there has been a revival of interest in the use of reaction time (RT) to study cognitive processes. Perhaps the most influential use of RT procedures has come from the memory search work of Sternberg (1969b). In a typical experiment, *S*s are shown a small set of items and asked to memorize them. They are then shown a single item (the test stimulus) and asked to pull one response lever if the test stimulus was in the memorized set and another if not. The size of the memorized set (M) is varied and the average RT for each set size is obtained. Sternberg (1966, 1967) has found in this situation that RT is a linear function of M and that the functions for positive and negative responses are parallel. Sternberg (1969a) has offered a theory of processing in this task. The theory is cast in the form of a description of a sequence of stages. At Stage 1 (Stimulus Preprocessing) a representation of the test stimulus is formed. At Stage 2 (Stimulus Comparison), this representation is compared with each of the items in the target set one at a time. At Stage 3 (Register Checking), the outcome of Stage 2 is evaluated, e.g., whether or not a match has been encountered. At Stage 4 (Response Selection), an appropriate response is selected and, at Stage 5 (Response Execution), this response is executed.

Two aspects of this theory of stages should be emphasized. First, the stages described above are presumed to reflect not only separate classes of information-processing operations, but more strongly are presumed to occur in strict temporal order. That is, a subsequent stage is not initiated until the previous stage has been completed. In the present research, the question is raised whether Stimulus Comparison and Response Selection necessarily reflect nonoverlapping stages of processing. An alternative would be a conception whereby Response Selection can be initiated while comparison operations are being carried out. Thus, at least in part, these stages could occur in parallel. Second, the operations at the Stimulus Comparison stage are carried out exhaustively. That is, the test stimulus is compared with all of the target set items regardless of the outcome of any early comparison. Sternberg's (1969a) explanation for exhaustive comparisons is reflected in the operations postulated at

---

[1] Experiment I was supported by Grant 22022 from Oakland University. Experiment II was supported by Grant MH-18702-01 from the National Institute of Mental Health. The authors wish to thank Gary Klein and D. James Dooling for their help.

[2] Requests for reprints should be sent to Joseph Dumas, Department of Psychology, Oakland University, Rochester, Michigan 48063.

Stage 3, the Register Checking stage. It is not until this stage that *S* receives and makes use of the outcome of the operations at Stage 2. Recently Checkosky (1971) and Dumas (1972) have reported evidence that indicates that the memory comparison process is a self-terminating dimension-by-dimension examination of the memory set. The present research also questions the exhaustiveness of the comparison process and indicates that *S*s can make use of feedback from early comparison operations before all potentially informative comparisons have been completed.

The major independent variable in the present research is the probability that a response (YES vs. NO) will be correct given the presence of particular attributes in the test stimulus. Consider the set of stimuli used in Exp. I (see Fig. 1). In the positive set, the shape dimension has three attributes (circle, square, and triangle) and clock position two attributes (12 o'clock and 4 o'clock). In a block of 48 trials, half of which required a YES response, each positive set pattern occurred eight times. In the negative set, there were two groups of patterns. The Clock NOs which share the clock position attributes with the positive set occurred three times each, while the Shape NOs which share the shapes occurred twice each. The probability that the YES response is correct given a 12 o'clock attribute in the test stimulus was .72 (16 out of 22). On the other hand, the 4 o'clock attribute gives relatively little information about which response is most likely (.57 vs. .43). If *S*s can use feedback from early feature tests, then, when a 12 o'clock attribute is in the test stimulus, *S* might begin selecting the YES response before completing his examination of the shape dimension. If this were the case, then RTs to 12 o'clock positive set patterns would be faster than to 4 o'clock positive set patterns. The responses to 12 o'clock Clock NO patterns, however, should be slower than to 4 o'clock Clock NO patterns since, upon examination of the shape, *S* would have to inhibit the processing of the positive response and begin the selection of the negative response. Experiment I tested this possibility.

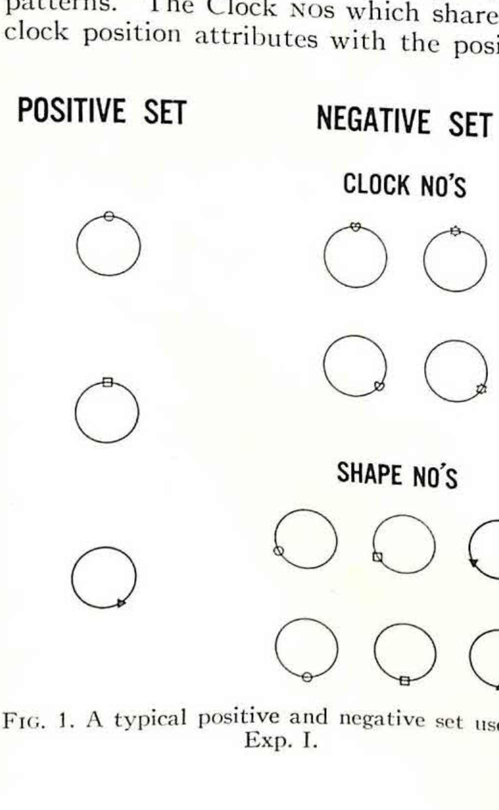


FIG. 1. A typical positive and negative set used in Exp. I.

## EXPERIMENT I

### *Method*

*Design.*—Each *S* was presented with a single positive set of three, two-dimensional patterns. In each positive set, the shape dimension contained three attributes while clock position had only two, one of which, the high-probability (HP) attribute, occurred with two patterns and the other, the low-probability (LP) attribute, occurred only once. The Clock NOs contained four patterns which shared the positive set clock positions; two shared the HP attribute while the other two shared the LP attribute. The Shape NOs consisted of six patterns that shared a positive set shape. Seven pairs of positive sets were used. In each pair, the shapes and clock position attributes were the same, but the clock position that was the HP attribute in one set was the LP attribute in the other. In all cases, clock positions that were 4 hr. apart were used in the positive set. The HP and LP attributes were always clock positions.

*Subjects.*—The *S*s were 14 right-handed female Oakland University undergraduates who were paid for their participation.

*Apparatus.*—Stimuli were projected on a white screen by a Kodak Carousel KT-800 projector. The *S*s' index fingers rested on pastic keys which, when pressed, activated a microswitch. Reaction times were measured by a Lafayette Instrument Co. digital display stop clock accurate to 1/100 sec. The clock was initiated with the onset of the slide and stopped when *S* pressed a key. The sequence of ready signal, presentation, and reset was controlled by an electronic timer. Reaction time and response (YES vs. NO) were recorded manually by *E*.

*Stimuli.*—Each pattern consisted of a geometric shape at a position on the perimeter of a circle. The five shapes were a circle, square, triangle, heart, and a six-pointed star. The shapes could appear on the circle at any of the even-numbered clock positions. Both the circle and the shapes were drawn in black on a white sheet of paper and Xeroxed onto transparent Mylar paper. The resulting transparency was then mounted on a Carousel slide mount. When projected on a white screen 4 ft. before *S*, a slide occupied 13 × 9 in. The circle representing the clock had a diameter of 7 in. when projected. The shapes were approximately 1 in. across.

*Procedure.*—Before the trials each day, each *S* was shown the positive and negative sets and given instructions stressing accuracy relative to speed. The sequence on each trial consisted of a 1-sec. warning buzzer followed 1 sec. later by a 2-sec. stimulus presentation. The interstimulus interval was 5 sec. After each response, *S*s were given feedback about accuracy, and if the response was correct they were also told how fast the response was. Each session was made up of two blocks of 48 trials each, with half of the stimuli requiring the positive response. The sequence of slides was randomized with the restriction that no more than three correct positive or three correct negative responses could occur in a run. Furthermore, a pattern could never follow itself. For the data analyses, the RTs for the first day were excluded to permit *S*s to stabilize their RTs. The basic score for the analysis of the RT data was a mean for each *S* for each block of the RTs to each condition. A count of the total response errors was tabulated separately for positive and negative responses for each condition to provide a relative frequency of error (RFE).

### *Results*

The mean RT for each stimulus type and response averaged over days, along with the RFE, appears in Table 1. The RTs for the YES responses to the two patterns sharing the HP clock position averaged 68 msec. faster than the RTs to the patterns having the LP clock position. However, for the NO responses, HP patterns averaged 20 msec. slower than the LP patterns. Statistically these data produced a significant main effect of response type, $F\ (1,\ 13) = 82.6,\ p < .01$, as well as a significant Attribute Probability (HP Attribute vs. LP Attribute) × Response Type interaction, $F\ (1,\ 13) = 4.96,\ p < .05$. Further, as can be seen in Table 1, the error data mirror the RTs, with the HP clock position YES responses producing a lower error percentage than the LP clock position YES responses, while the reverse is true for the NO responses. Finally, *S*s were faster on Day 3 than on Day 2, $F\ (1,\ 13) = 32.3,\ p < .001$. However, there were no significant interactions with days.

TABLE 1

AVERAGE CORRECT REACTION TIME (RT) AND RELATIVE FREQUENCY OF ERROR (RFE) TO THE HIGH-PROBABILITY (HP) AND LOW-PROBABILITY (LP) ATTRIBUTE PATTERNS FOR BOTH YES AND CLOCK NO RESPONSES AND TO THE SHAPE NO PATTERNS

| Measure | Response | | | | |
|---|---|---|---|---|---|
| | YES | | Clock NO | | Shape NO |
| | HP | LP | HP | LP | |
| RT | 801 | 869 | 943 | 923 | 872 |
| RFE | .019 | .032 | .074 | .006 | .015 |

## EXPERIMENT II

The Attribute Probability × Response Type interaction in Exp. I suggests that *S*s receive and make use of feedback from early feature tests in order to initiate response selection *before* all potentially informative feature tests are completed. If the additional feature tests are completed before the response is executed, *S* uses feedback from these feature tests either to continue selecting and executing the response that had been initiated (resulting in a fast RT) or he inhibits this response and begins selecting the alternate response (resulting in a slow RT). However, on some of the trials, response execution is completed before the remaining feature tests and *S* produces an error.

This analysis would seem to require that the RTs to Shape NO patterns be equal to RTs to the LP-attribute Clock NO patterns. Since the LP attribute gives little information about response probability and, since both Shape NO patterns and Clock NO patterns each share one attribute with the positive set, their RTs should be equal. However, the LP-attribute Clock NO RTs were 51 msec. slower than the Shape NO RTs. This finding could be accounted for by assuming that *S*s in Exp. I did not randomly select which dimension to examine first. If they were biased toward looking at the clock position dimension first, their RTs to Shape NO patterns would be faster than their RTs to the LP-attribute Clock NO patterns.

Experiment II was conducted to replicate the results of Exp. I while providing better control over some possible extraneous variables. While there is no a priori basis for estimating the times for the proposed process, there are a number of difficulties in using the times found in Exp. I. First of all, there were some individual differences in the Attribute Probability × Response Type interaction. For four of the 14 *S*s the HP Clock NO RTs were not slower than the LP Clock NO RTs. Second, the clock position dimension always contained the HP and LP attributes. Thus some of the results found in Exp. I may be specific to that dimension. Third, because of the makeup of the stimulus sets in Exp. I, the HP attribute occurred more often than any other attribute. This leaves the possibility open that effects due to attribute probability occur only in the situation where the most frequently appearing attribute is mapped unequally to the response set throughout practice.

### *Method*

*Design.*—Each *S* was presented with four conditions on each of 6 days. In two of the conditions, the positive set was similar to that of Exp. I. The two conditions differed only in which dimension (shape or color) contained the HP and LP attributes. In the color condition the color dimension contained the HP and LP attributes (see Fig. 2), while in the shape condition it was the shape dimension (e.g., a red square, a green square, and a blue circle). Notice in Fig. 2 that there were three categories of NO response based on their similarity to the positive set. Each NO-2 pattern shared two attributes with the positive set. One of the NO-2 patterns shared the HP attribute with the positive set, while the other shared the LP attribute. Each of the NO-1 patterns shared one value with the positive set, in one case the color and in the other the shape. Finally, neither of the NO-0 patterns shared an attribute with the positive set. Since all nine of these patterns were presented an equal number of times in a block of trials, the HP attribute occurred the same number of times as the LP attribute.

The other two conditions were included for two reasons. First, as a replication of two of the conditions used by Checkosky (1971) and Dumas (1972) and, second, as a test of the suggestion in Exp. I that *S* biases the order of examination of dimensions when attribute probability is varied. Both of these conditions contained negative sets similar to that shown in Fig. 2. They differed from the color and shape conditions only in the makeup of their positive sets. In the six-attribute condition, there were three patterns in the positive set, none of which had an attribute in common (e.g., a red circle, a green square, and a blue triangle.) In the four-attribute condition, there were only two patterns in the positive set.

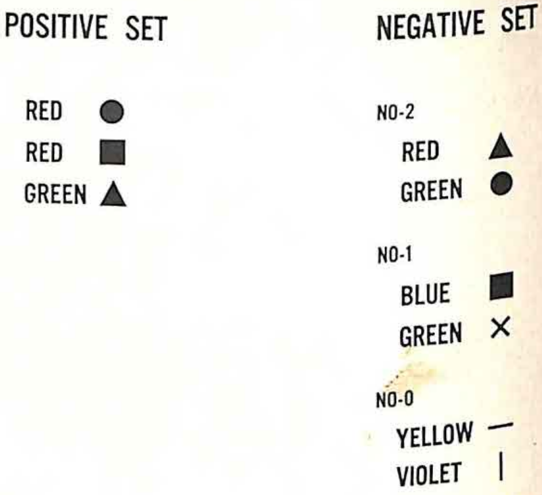


FIG. 2. A typical positive and negative set used in Exp. II.

*Subjects.*—The *S*s, two males and six females, were Lake Forest undergraduates who were paid for their participation.

*Apparatus.*—Stimuli were generated by an in-line display (Lehigh Valley Standard Pattern #696). The pre- and postexposure fields as well as the background field were uniformly black. Reaction times were measured by a stop clock accurate to 1/100 sec. The clock was initiated with the onset of the pattern and stopped when *S* pressed a key.

*Stimuli.*—The stimuli consisted of white outlines in colored backgrounds. The colors were background circles 1 in. in diameter with hues of red, green, amber, violet, blue and pink. The six shapes were outlines of a circle, square, triangle, a horizontal line segment, a vertical line segment and an "X." These shapes were clear and were centered inside the colored circles.

*Procedure.*—Each *S* participated in four conditions on each of six sessions separated by at least 1 day but not more than 2. The order of the conditions was counterbalanced over *S*s on each day. Each condition consisted of 24 positive set stimuli and 48 negative set stimuli. Every shape and color appeared equally often in the positive set for each condition over the last 4 days of practice. In addition, each *S*'s positive and negative set for each condition was different for each of the last 4 days. Only the data from the last 4 days of practice was used in the analyses.

*Results*

The results from the conditions which replicate those of Exp. I will be reported first. The mean RTs and RFE to the HP- and LP-attribute YES and NO-2 responses along with the NO-1 and NO-0 responses appear in Table 2. These responses have been collapsed over days, and the shape and color conditions have been combined since these effects were not significant and did not interact with any of the other variables. As in Exp. I, RTs to the HP-attribute positive set patterns were faster than RTs to the LP-attribute positive set patterns. However, RTs to the HP-attribute NO-2 patterns were slower. As in Exp. I, the RFE mirrored the pattern of RTs. An analysis of variance computed on the YES and NO-2 responses showed that both response type, $F\ (1, 7) = 47.34$, $p < .01$, and the Attribute Probability × Response Type interaction, $F\ (1, 7) = 62.40$, $p < .01$, were significant.

Table 3 presents the mean RTs and RFE for the four- and six-attribute conditions. Statistically, there was a significant effect of type of NO (NO-2 vs. NO-1 vs. NO-0), $F\ (2, 14) = 84.83$, and a significant memory load effect (two patterns vs. three patterns in the positive set) for both YES, $F\ (1, 7) = 13.93$, and NO responses, $F\ (1, 7) = 29.43$, all $p$s $< .01$. However, there were no interactions with memory load. In addition to replicating earlier findings, the results from these conditons can be used to test the assumption that *S* randomly chooses which dimension of the test stimulus to examine first. In all of the conditions, one NO-1 pattern shared a color with the positive set and the other shared a shape. In both the four- and six-attribute conditions, there was no significant difference between RTs to these two types of patterns ($F$ values less than 1.0). This finding is consistent with the assumption that for these conditions, *S*s randomly choose which dimension to examine first. However, in the color and shape conditions, there was a difference between the two NO-1 patterns. In both cases, RTs to the NO-1 pattern which did not share an attribute with the dimension in the positive set that contained the HP and LP attributes (in Fig. 2, the blue square) were faster than RTs to the other pattern. Thus the average RT to NO-1 patterns that shared the dimension containing the HP attribute was 473 msec., while the average RT to patterns containing an attribute from the other dimension was 493 msec., $F\ (1, 7) = 9.27$, $p < .05$. This effect did not interact with the color and shape conditions. This finding suggests that *S*s biased the order in which the dimensions were examined.

TABLE 2

AVERAGE CORRECT REACTION TIME (RT) AND RELATIVE FREQUENCY OF ERROR (RFE) TO THE HIGH-PROBABILITY (HP) AND LOW-PROBABILITY (LP) ATTRIBUTE PATTERNS FOR BOTH YES AND NO-2 RESPONSES AND TO THE NO-1 AND NO-0 PATTERNS

| Measure | Response | | | | | |
|---|---|---|---|---|---|---|
| | YES | | NO-2 | | NO-1 | NO-0 |
| | HP | LP | HP | LP | | |
| RT | 499 | 543 | 647 | 574 | 473 | 457 |
| RFE | .019 | .057 | .078 | .023 | .002 | .001 |

TABLE 3

AVERAGE CORRECT REACTION TIME (RT) AND RELATIVE FREQUENCY OF ERROR (RFE) FOR THE YES, NO-2, NO-1, AND NO-0 RESPONSES FOR THE FOUR-ATTRIBUTE (FA) AND SIX-ATTRIBUTE (SA) CONDITIONS

| Cond. | Response | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | YES | | NO-2 | | NO-1 | | NO-0 | |
| | RT | RFE | RT | RFE | RT | RFE | RT | RFE |
| FA | 482 | .060 | 575 | .092 | 462 | .008 | 420 | .001 |
| SA | 536 | .070 | 623 | .095 | 500 | .005 | 476 | .002 |

## DISCUSSION

Experiment II, while controlling overall stimulus probability and varying attribute probability with two new dimensions, replicated the results of Exp. I. The results of these two studies demonstrate that when an attribute of a dimension is unevenly mapped to the response set and when that attribute is contained in a pattern which requires the more probable response, RTs are fast; but when it is contained in a pattern which requires the less probable response, RTs are slowed. Furthermore, attribute probability affected the order in which dimensions were examined.

Two conclusions follow from these results. First of all, in a memory search task the initiation of Response Selection does not have to wait for the final results of the matching process. This is inconsistent with the rationale for exhaustive scanning, namely, that the outcome of the matching process is unknown until a match register is checked. In the present experiments, *S* apparently has information from early comparisons available before all the attributes are compared and uses this information when he can predict with high likelihood the probability of a positive response. In experiments which use stimuli which cannot be easily analyzed into separable attributes (Garner & Felfoldy, 1970), *S* has no partial information which would allow the overlapping with response selection to manifest itself. Because the stimuli used in the present experiments have separable attributes, they allow *S* to use partial information to make some estimate of response probabilities. Furthermore, there are two aspects of the present data which indicate that stimuli with separable attributes fall within the boundary conditions for testing the serial exhaustive model. The memory load differences found in Exp. II and by Checkosky (1971) and Dumas (1972) are similar to those found for stimuli with nonseparable attributes (e.g., Sternberg, 1967), and these memory load differences for the two types of stimuli are independent of response type.

Second, the matching of the test stimulus with the memory set is a self-terminating dimension-by-dimension feature-testing process. Consistent with the findings of Checkosky (1971) and Dumas (1972), the differences between the types of NO responses in Exp. II indicate that the comparison process is sensitive to the similarity of the test stimulus features to the memory set features. The utilization of early matching information found in the present experiments is consistent with a self-terminating process, since self-termination requires that matching information become available before the memory set is exhaustively examined.

## REFERENCES

CHECKOSKY, S. Speeded classification of multidimensional stimuli. *Journal of Experimental Psychology*, 1971, **87**, 313–318.

DUMAS, J. Scanning memory for multidimensional stimuli with extended practice. *Perception & Psychophysics*, 1972, in press.

GARNER, W., & FELFOLDY, G. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, **1**, 225–241.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–564.

STERNBERG, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

STERNBERG, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica*, 1969, 30, 276–315. (a)

STERNBERG, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457. (b)

(Received September 22, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 333–337

# FALSE RECOGNITION AS A FUNCTION OF ENCODING DIMENSION AND LAG[1]

GLEN A. RASER[2]

*University of Michigan*

The effects of acoustic and orthographic similarity on false recognition responses were examined in two experiments. Word pairs having either a high degree of similarity on both dimensions, on the acoustic dimension only, on the orthographic dimension only, or on neither dimension were presented in a continuous recognition memory task at various lags. False-alarm increases as a function of similarity were found for both dimensions. The effect was greater with the orthographic dimension but only because some *S*s made only orthographic false alarms while others made both acoustic and orthographic false alarms. Both orthographic and acoustic similarity produced similarly shaped lag functions with false alarms rising rapidly over the shortest lags and remaining relatively constant thereafter.

Several recent conceptualizations of human memory have emphasized the role of attributes in the storage and retrieval of information (e.g., Norman & Rumelhart, 1970; Underwood, 1969). According to these notions, an event is stored in memory simply as a list of attribute values along a set of dimensions relevant to that event. Even when limited to verbal stimuli there is a sizable number of effective encoding dimensions (Wickens, 1970). In the following experiments, two related dimensions of individual words, the acoustic and the orthographic dimensions, are compared with regard to their effectiveness as encoding dimensions at various lags in a continuous recognition memory task. These two dimensions were chosen for study because acoustic properties of words have been shown to affect performance in a variety of memory tasks (Shulman, 1971). In general, however, the acoustic and orthographic properties of words are highly confounded so that it is not always clear whether particular results are due to acoustic or orthographic effects.

The continuous recognition memory paradigm was used in these experiments for two major reasons. First, the paradigm offers a relatively direct method of determining when a particular encoding dimension is being used. If, for example, a word is presented for recognition and then followed at some later point by another word similar to the first along some dimension and *S* indicates that he recognizes the second word (makes a false alarm) at an above-chance rate, then it may be inferred that the dimension on which the two words were similar was used to encode the two events. The second important feature of the paradigm is that the encoding process for a particular item may be assumed to remain constant regardless of the lag at which that item is tested. This is true because *S* has no way of knowing at the time of the first presentation of an event when that event will be tested.

Considering the evidence suggesting that short-term memory (STM) may be primarily acoustic in nature (Baddeley, 1966; Kintsch & Buschke, 1969), the false recognition rate based on acoustic similarity should be higher at shorter lags than at longer ones. Thus the lag manipulation was introduced to provide a more detailed comparison of the acoustic and orthographic dimensions.

---

[1] These experiments were conducted while the author was the recipient of a National Institute of Mental Health Postdoctoral Fellowship and were supported, in part, by United States Department of Health, Education, and Welfare Grant OEG-0-9-320447-4194 (010), to James G. Greeno.

[2] Requests for reprints should be sent to Glen A. Raser, Human Performance Center, 330 Packard Rd., University of Michigan, Ann Arbor, Michigan 48104.

## METHOD

### Design

*Experiment I.*—A basic continuous recognition memory list included four sets of word pairs, with each set having a particular similarity relationship between the members of a pair within the set, arranged at lags of either 1, 2, 4, 10, or 30 (0, 1, 3, 9, or 29 intervening items, respectively, between the occurrence of the first and second members of a related pair). The four similarity relationships included two levels of acoustic similarity (high or low) factorially crossed with two levels of orthographic similarity (high or low). Each of the four pair types was used equally often at each of the five lags. Within a single word pair, there was a first presentation (a word in the list without a preceding similar word) which could be used to estimate the usual false-alarm rate and a second presentation (a word preceded by a similar word) which could be used to estimate the similarity effect on false-alarm rate. Thus the design was a $2 \times 2 \times 5 \times 2$ factorial with the respective factors being acoustic similarity, orthographic similarity, lag, and presentation.

*Experiment II.*—The design was similar to that of Exp. I, differing only in that a list factor with 10 levels was added. The 10 lists resulted from using each word pair within a set at each of the five lags and reversing the presentation order within a single pair at each lag condition.

### Stimulus Materials

A pair of words was considered of high acoustic similarity only if their pronunciation was identical (i.e., homophone pairs) and of high orthographic similarity if they had the same length and differed only by one letter other than the initial or final letter. A pair of words was considered of low acoustic similarity if they were not pronounced identically and of low orthographic similarity if they were not of the same length and if one pair member contained at least two letters not contained in the other pair member. Thus, the four similarity relationships (with examples) were as follows: high orthographic and high acoustic (HO HA), FOUL-FOWL; high orthographic and low acoustic (HO HA), SOUR-SLUR; low orthographic and high acoustic (LO HA), CLAWS-CLAUSE; and low orthographic and low acoustic (LO LA), CLOUD-CLEAVE. HO HA and LO HA word-pair sets were selected from a larger set of homophones collected in an earlier study (Raser, 1970), and HO LA and LO LA word-pair sets were selected from Thorndike and Lorge.

Experiment I included 40 pairs within each of the four sets along with 80 repeated words and a primacy buffer of 30 repeated words in the complete recognition list. Four different lists were constructed by making four different random assignments of word pairs within a set to particular lag conditions.

In Exp. II, additional care was taken in selecting word pairs so that the four pair types were closely matched for word frequency values. Also LO LA pairs were selected to match LO HA pairs for word length and initial letter. A complete recognition list included 25 pairs within each of the four sets along with 100 repeated words and a primacy buffer of 30 repeated words. In both experiments, the same lag values were used for repeated words as for the various similarity conditions. Ten different lists were constructed by counterbalancing specific pairs within a set across all levels of lag and presentation.

### Subjects

A total of 66 students, 16 in Exp. I and 50 in Exp. II, from the Human Performance Center *S* pool at the University of Michigan served as *S*s and were paid $1.50 for their participation. The *S*s were run in groups of 2–5. In Exp. II, 5 *S*s were assigned to each of the 10 lists in the order of their appearance at the laboratory.

### Procedure

*Experiment I.*—The lists were constructed and presented by computer at individual display stations. Each word appeared for 2.5 sec. during which time *S* made a yes or no response by pressing the appropriately labeled button located on either side of the display screen. The *S*s were instructed to give a yes response if they thought they had seen the word previously in the list and a no response if they thought it was a new word.

*Experiment II.*—The same apparatus for presentation and responding was used in Exp. II. Each word appeared on the screen for 5 sec. with a yes or no response required during the first 2.5 sec. During the remaining 2.5-sec. period, *S* gave a certainty rating on a 3-point scale of the correctness of his response by typing in the appropriate number on the keyboard below the screen. Response latencies were also recorded in both experiments for the recognition response.

## RESULTS AND DISCUSSION

The main dependent variable in both experiments was the number of false alarms per condition with a possible range of 0–8 in Exp. I and 0–5 in Exp. II. Analysis of variance was used in both cases and revealed quite similar patterns of results. Significant main effects were found for orthographic similarity, $F\ (1, 15) = 19.23$, $p < .01$ in Exp. I, $F\ (1, 40) = 83.69$, $p < .01$ in Exp. II; acoustic similarity, $F\ (1, 15) = 8.75$, $p < .01$ in Exp. I, $F\ (1, 40) = 43.91$, $p < .01$ in Exp. II; lag, $F\ (4, 60) = 3.24$, $p < .05$ in Exp. I, $F\ (4, 160) = 6.27$, $p < .01$ in Exp II;

and presentation, $F$ (1, 15) = 26.74, $p < .01$ in Exp. I, $F$ (1, 40) = 123.24, $p < .01$ in Exp. II. Consistently significant two-way interactions were found when the presentation factor was involved. These included Orthographic Similarity × Presentation, $F$ (1, 15) = 26.80, $p < .01$ in Exp. I, $F$ (1, 40) = 126.17, $p < .01$ in Exp. II; Acoustic Similarity × Presentation, $F$ (1, 15) = 4.60, $p < .05$ in Exp. I, $F$ (1, 40) = 11.26, $p < .01$ in Exp. II; and Lag × Presentation, $F$ (4, 60) = 3.11, $p < .05$ in Exp. I, $F$ (4, 160) = 8.00, $p < .01$ in Exp. II. The only other consistently significant higher order interactions were Orthographic × Acoustic Similarity, $F$ (1, 15) = 5.71, $p < .05$ in Exp. I, $F$ (1, 40) = 5.00, $p < .05$ in Exp. II and Orthographic Similarity × Acoustic Similarity × Presentation, $F$ (1, 15) = 16.26, $p < .01$ in Exp. I, $F$ (1, 40) = 5.66, $p < .05$ in Exp. II. Experiment II also included a lists factor which did not appear as a significant main effect.

The results can be best interpreted by considering the interactions. First, both the orthographic and acoustic dimensions served as relevant encoding dimensions in this task as evidenced by the significant interactions of the similarity factors with the presentation factor. The exact form of these interactions can be seen in Fig. 1, which shows the interaction of all three factors, by comparing the solid lines in each graph with the dashed lines for the orthographic effect and the closed circles with the open circles for the acoustic effect. It is also apparent that the orthographic dimension was much more effective in producing false alarms than the acoustic dimension. This finding is of particular importance because of the emphasis placed on acoustic effects by some STM memory theories. It may be that many demonstrations of apparent acoustic effects in memory should more properly be labeled orthographic effects because of the high degree of confounding usually existing between the two dimensions.

The lag functions for the particular similarity relationships examined in these studies are also of interest and are shown in Fig. 2 along with the hit rates for repeated words. In general, the false-alarm level increases as lag is increased from 1 to 2 and remains relatively constant thereafter. The lag factor failed to show any consistent interactions with the similarity factors. In short, there is no evidence for either acoustic or orthographic effects being confined to any one retention interval or set of retention intervals.

Two additional analyses of variance were performed using the latency data from each experiment. Those word pairs which received a no response on both the first and second presentation were selected and the mean differences in response latencies (second presentation latency minus first presentation latency) for each $S$ for each Lag × Pair Type Condition in Exp. I and over a set of five $S$s for each Lag × Pair Type Condition in Exp. II were used as the dependent variables. Positive latency dif-

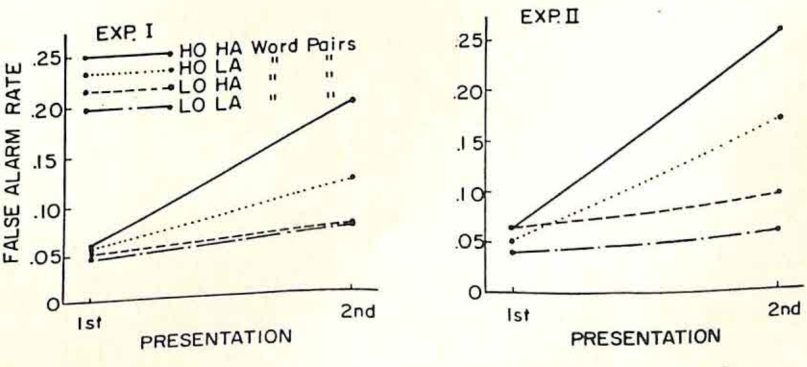


FIG. 1. False-alarm rate as a function of pair type and presentation for Exp. I and II.

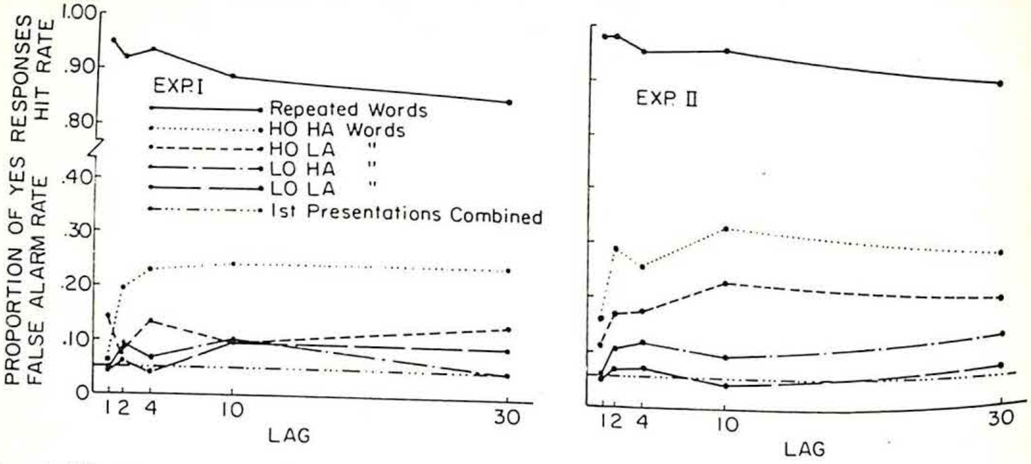
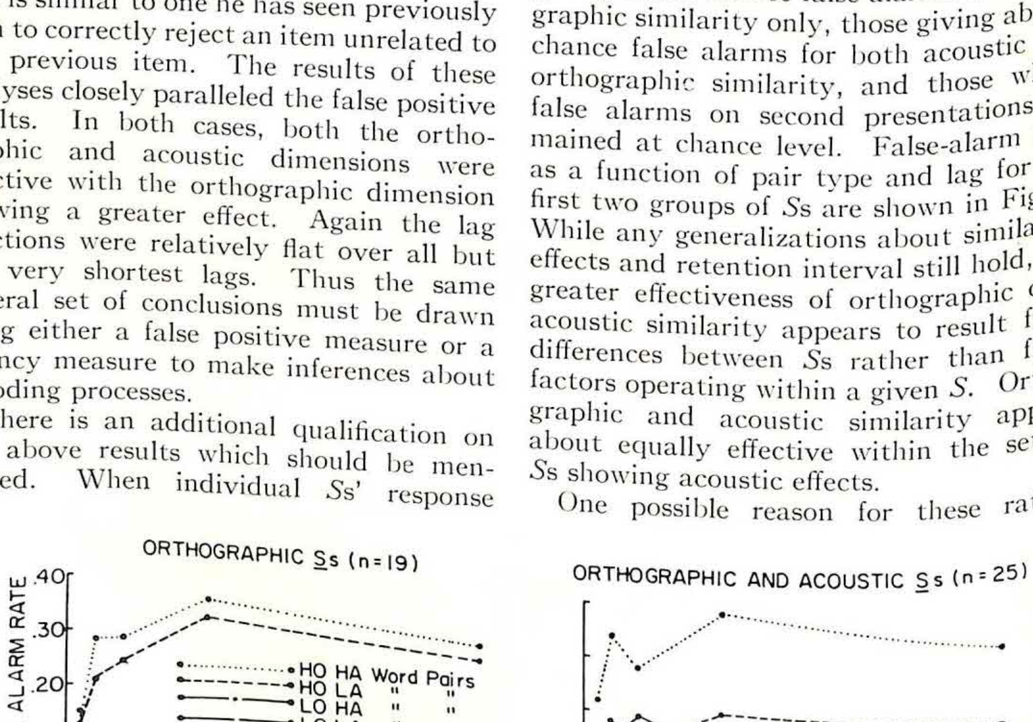


FIG. 2. Hit rate for repeated words and false-alarm rates for the first and second presentations of the four word-pair types as a function of lag for Exp. I and II.

ferences are interpretable as reflecting the effectiveness of a particular encoding dimension as it is reasonable to assume that *S* will take longer to correctly reject an item that is similar to one he has seen previously than to correctly reject an item unrelated to any previous item. The results of these analyses closely paralleled the false positive results. In both cases, both the orthographic and acoustic dimensions were effective with the orthographic dimension showing a greater effect. Again the lag functions were relatively flat over all but the very shortest lags. Thus the same general set of conclusions must be drawn using either a false positive measure or a latency measure to make inferences about encoding processes.

There is an additional qualification on the above results which should be mentioned. When individual *S*s' response patterns for Exp. II were examined, it soon became apparent that *S*s could be easily categorized into three groups, those giving above-chance false alarms for orthographic similarity only, those giving above-chance false alarms for both acoustic and orthographic similarity, and those whose false alarms on second presentations remained at chance level. False-alarm rate as a function of pair type and lag for the first two groups of *S*s are shown in Fig. 3. While any generalizations about similarity effects and retention interval still hold, the greater effectiveness of orthographic over acoustic similarity appears to result from differences between *S*s rather than from factors operating within a given *S*. Orthographic and acoustic similarity appear about equally effective within the set of *S*s showing acoustic effects.

One possible reason for these rather


ORTHOGRAPHIC Ss (n=19)

ORTHOGRAPHIC AND ACOUSTIC Ss (n=25)

FIG. 3. False-alarm rates for second presentations as a function of pair type and lag for orthographic *S*s and for orthographic and acoustic *S*s from Exp. II.

striking individual differences is that those *S*s showing both acoustic and orthographic effects may have been repeating the words to themsleves and thus providing both visual and acoustic input while the remaining *S*s may have avoided such activity. Another possibility is that some *S*s may simply prefer using a visual memory code. However, there is no justification for prefering one hypothesis over the other given the data available from this study.

The above obtained results are partially consistent with those from a study by Bregman (1968) which compared four kinds of recall cues in a task similar to the continuous recognition memory task. Bregman presented lists of words with either contiguity, phonetic, graphic, or semantic recall cues at lags ranging from 1 to 96 after presentation of the target words. His phonemic and graphic cue conditions are of particular interest. In both cases, the effectiveness of the cue in producing recall declined rapidly over the shortest few lags with a greatly reduced rate of decline thereafter. His retention curves for recall using phonemic and graphic cues were virtually indistinguishable. Thus while there were no interactions with lag in his study, there was also no evidence to indicate any greater effectiveness of one cue type over the other.

In summary, the above results provide evidence for the use of orthographic features, to an even greater extent than acoustic features were used, in a continuous recognition memory task with visual presentation. This result was consistent across a wide range of lag intervals and was obtained with variation in presentation rate and with different proportions of repeated items in the recognition list.

## REFERENCES

BADDELEY, A. D. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362–365.

BREGMAN, A. S. Forgetting curves with semantic, phonetic, graphic, and contiguity cues. *Journal of Experimental Psychology*, 1968, **78**, 539–546.

KINTSCH, W., & BUSCHKE, H. Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, 1969, **80**, 403–407.

NORMAN, D. A., & RUMELHART, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

RASER, G. A. Encoding processes in recognition memory. Unpublished doctoral dissertation, Iowa State University, 1970.

SHULMAN, H. G. Similarity effects in short-term memory. *Psychological Bulletin*, 1971, **75**, 399–415.

UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received September 29, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 338–342

# MODIFICATION OF PERCEPTION BY CLASSICAL CONDITIONING PROCEDURES[1]

C. RICHARD CHAPMAN AND BEN W. FEATHER[2]

*Duke University Medical Center*

Perceptual sensitivity ($d'$) and response bias ($C_k$) were estimated in a task where *S*s were required to recognize a brief visual stimulus. The stimulus (CS) was then displayed 5 sec. before an electric shock (UCS) in a series of classical conditioning trials. Repetition of the stimulus recognition task demonstrated significant $d'$ increases and significant $C_k$ decreases. No such changes were evident in controls who received no conditioning. The relevance of this finding to perceptual defense is noted.

The ability of an individual to perceive a brief visual stimulus having strong emotional associations has been studied extensively. The literature on perceptual defense has examined increases in recognition thresholds for stimuli which are threatening or taboo compared to those for neutral stimuli. In a review of this literature, Natsoulas (1965) noted that response variables are extremely important in the defense phenomenon. He suggested that perceptual defense may reflect the effects of emotionally charged stimuli on the response system while the perceptual system remains unaffected.

Goldiamond (1958) reported that perceptual vigilance or sensitization is also observed for certain kinds of threatening word stimuli; i.e., lower thresholds are evident for the recognition of affective words than for the recognition of neutral words. Examining perceptual defense and vigilance in light of the psychophysical methodology employed in this literature, he concluded that most of the perceptual defense or vigilance research has failed to provide information about perceptual processes, since response variables significantly affect thresholds. After evaluating the assumptions of threshold theory in contrast to the alternative methodology provided by signal detection theory, Goldiamond suggested that the threshold approach is an inappropriate psychophysical tool for the study of these phenomena.

Signal detection theory (Green & Swets, 1966) provides a model which construes perception as both a sensory experience and a decision-making response involving the processing of the sensory experience. The methodology associated with this model yields two numerical indices for perception in place of the single threshold estimate. The first, $d'$, is a relatively pure estimate of sensitivity to the physical stimulation. The second measure, $C_k$, represents the perceiver's response bias, i.e., his willingness to report that he has seen the stimulus.

Two studies have examined the perceptual defense and perceptual vigilance phenomena using detection theory methodology. Van Egeren (1968) presented neutral and affect-arousing words to *S*s tachistoscopically. No differences in $d'$ were evident when recognition of neutral words was compared with recognition of affect-arousing words. Significant differences were evident, however, in response bias for the two types of words such that *S*s were less willing to report the recognition of affect-arousing stimuli. These results demonstrate a defense effect which is purely a response variable phenomenon.

Dorfman (1967) presented taboo words tachistoscopically to *S*s, varying the a

---

[1] This investigation was supported in part by Research Training Grant 5 T01H00164 of the National Institute of Child Health and Human Development, by a Public Health Service Career Development Award 5 K02 19523-08, and by National Institute of Mental Health Grant 5 R01 MH 11549-05.

[2] Requests for reprints should be sent to C. Richard Chapman, Department of Anesthesiology, Mail Stop RC-40, University of Washington School of Medicine, Seattle, Washington 98105.

priori probability of the taboo words. He observed that *d′* was higher for taboo words than for neutral words, a perceptual vigilance phenomenon, and that response bias varied with a priori probability of the taboo words. On the basis of the Zajonc and Dorfman (1965) observation that conditioned fear increases the perceived intensity of a stimulus, Dorfman hypothesized that conditioned fear was associated with taboo words and that this fear had increased the discriminability of the taboo words his *S*s viewed.

The present experiment is a further investigation of the perceptual processes basic to the recognition of threatening stimuli. In this study *S*s perceived a stimulus that had taken on affective qualities through classical conditioning procedures in which a neutral geometrical figure (CS) was followed by an electric shock (UCS). The effects of the conditioning process on perceptual sensitivity and on response bias were examined.

## METHOD

*Subjects.*—Thirty-six 18–25-yr.-old men from the Duke University community served as paid volunteers, receiving $5.00 each.

*Apparatus.*—A Kodak (Model 860) Carousel projected through the acoustical glass window of a sound-proof chamber either a random visual noise slide or a slide superimposing a triangle on the noise. The image appeared on a 91.5 × 61 cm. screen mounted approximately 183 cm. before *S*'s reclining chair. The speed of slide presentation was controlled by an Alphax shutter mounted on the telescoping projector lens and triggered by the action of a solenoid.

Shock electrodes were attached to the fourth digit of the nonpreferred hand of all *S*s receiving shock, the shock source being an autotransformer with a built-in stimulus isolation circuit.

The *S*s made responses to the stimuli by pushing one of four toggle switches mounted on a lap-held box, 30.5 × 6.3 × 6.3 cm. Two of the switches were labeled Triangle–Somewhat Certain, and Triangle–Certain. The others were labeled Blank–Somewhat Certain, and Blank–Certain.

A PDP-12 computer was used to control the random order of slide presentation, the triggering of the Alphax shutter, the delivery and duration of the shock, and to record and tally *S*s' responses to each slide.

*Experimental design.*—The *S*s were randomly assigned to three treatment groups of 12 *S*s each, an Experimental group and two Control groups. Measurements of perceptual processes were made both before and after treatment for each group. For the Experimental *S*s, treatment consisted of a conditioning session in which the triangle became the CS for a painful shock. A Random-Shock Control group experienced the same number of shocks but with no stimulus contingencies. A third group, No-Shock Control, received the same visual experiences as the preceding groups, but no shock was given or mentioned in the instructions. All *S*s were run individually.

*Procedure.*—The experiment was divided into three parts with an initial adjustment session preceding the first part.

The initial adjustment procedure served to (*a*) accommodate the difficulty of the task to the individual *S* and (*b*) to familiarize *S* with the experimental situation. The *S*s were instructed to view slide presentations in blocks of 20 trials and to report on each trial whether the triangle was present or absent, indicating the certainty of each decision. The background light level was adjusted until *S* responded within a specified performance range such that the probability of a correct detection was between $p = .6$ and $p = .8$. Blocks of 20 trials were repeated until *S* met this criterion.

Before Part 1, *S*s were told that a trial would begin with the offset of a white light 2 sec. before the slide presentation. The stimulus duration was .1 sec., and *S* was given 4 sec. in which to respond. At the end of the 4 sec., the onset of the white light marked the beginning of the intertrial interval (ITI), which lasted 4, 5, or 6 sec. on a random schedule. The offset of the light signaled the beginning of the next trial. If *S* failed to respond, the trial was repeated. Part 1 consisted of 150 trials with a 5-min. rest after 75 trials.

At the end of Part 1, electrodes were attached to *S*s in the Experimental and Random-Shock Control groups. The criterion for the level of shock was that it should be painful and above 2 ma. The mean shock level for the conditioning *S*s was 2.68 ma.; it was 2.93 ma. for the Random-Shock Control *S*s. Each *S* was told that 24 extended slide presentations would follow and that he would make no responses during these presentations. An equal number of triangles and blanks was presented, with the slides being in random order and of 6-sec. duration. Experimental *S*s received a 1-sec. duration shock 5 sec. after the onset of a triangle slide. These *S*s were told that the shock was contingent on the presentation of the triangle. Random-Shock Control *S*s received identical shocks, six occurred after triangle onset, and six occurred after blank onset, with the shock occurring 3, 5, or 7 sec. after the stimulus according to a random schedule. The third group, No-Shock Control, viewed the two slides at 6-sec. durations but received no shock. Randomly ordered ITIs of 19, 20, or 21 sec. were used for each group.

The third part of the experiment was identical to the first part with the exception that the Experimental *S*s and Random-Shock Control *S*s received four additional .5-sec. shocks at the usual intensity level. The Experimental *S*s were shocked .5 sec. after the brief presentation of the triangle at four points during Part 3. Random-Shock Control *S*s received two analogous shocks after the presentation of triangle slides and two shocks after the presentation of blank slides. Fourteen-second ITIs followed the shock trials. Two of the shock trials occurred during the first 75 trials of Part 3 and the remaining two occurred during the last 75 trials. No shocks were given to the third Control group.

## RESULTS

Responses to noise alone (N) and to the signal plus noise (SN) were analyzed separately for Part 1 and Part 3 of the experiment. The rating-scale methodology associated with signal detection theory (Hake & Rodwan, 1966) was used to estimate perceptual sensitivity ($d'$) and response bias ($C_k$). This process is best described as the generation of a receiver operating characteristic (ROC) for each *S*, i.e., a function relating correct identifications of SN, hits, to incorrect reports of N as SN, false alarms. In this study, each *S*'s rating category response tallies were cumulated from $k_4$ (Triangle–Certain) to $k_1$ (Blank–Certain) and then converted to probabilities. This procedure provided three sets of hit rate (HR) and false-alarm rate (FAR) values.

For simplicity, HR and FAR values were averaged across *S*s within groups, and ROCs were generated from these means to represent each of the three groups. By plotting HR and FAR for each category on double probability paper, each ROC was obtained as a linear function. Figure 1 presents the Part 1 and Part 3 ROC functions for each of the groups.

The nature of the ROC provides certain basic information about the underlying hypothetical probability distributions in the signal detection model (Green & Swets, 1966). Since each set of points fits a single straight line reasonably well, the distributions may be assumed to be normal. The slope of each function is a ratio of the standard deviations of the distributions; e.g., a slope of 1.0 means that the variances of the two distributions are equal. Table 1, which provides the slopes for each of the functions in Fig. 1, indicates that the slopes are consistently greater than 1.0.

While in most signal detection research it is assumed that the variances of the two distributions are equal, Green and Swets (1966) have noted that in many cases ROCs are asymmetrical and consistent with an alternative assumption that $\Delta M/\Delta\sigma = 4$ where $\Delta M$ is the difference between the means of the hypothetical distributions and $\Delta\sigma$ is the difference between the standard deviations. The $\Delta M/\Delta\sigma$ ratios in Table 1 suggest that these data fit the alternative assumption better than the more familiar equal variance model.

In order to evaluate the $d'$ differences for each group across Part 1 and Part 3, $d'$ values were estimated for each *S* according to the method introduced by Richards and Thornton (1970). Specifically, a two-parameter ($a$,$d'$) least-squares fit of *S*'s

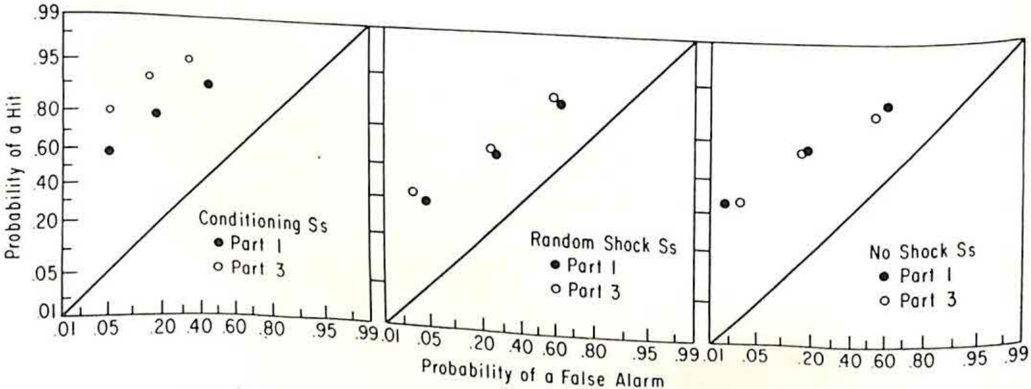


FIG. 1. Mean receiver operating characteristics for Part 1 and Part 3 of the experimental session.

TABLE 1

Mean Slopes and Mean-to-Sigma Ratios for Part 1 and Part 3

| Group | Measure | Part | |
|---|---|---|---|
| | | 1 | 3 |
| Experimental | Slope | 1.50 | 1.48 |
| | $\Delta M/\Delta\sigma$ | 4.15 | 5.99 |
| Random-Shock Control | Slope | 1.27 | 1.26 |
| | $\Delta M/\Delta\sigma$ | 4.67 | 6.26 |
| No-Shock Control | Slope | 1.59 | 1.46 |
| | $\Delta M/\Delta\sigma$ | 3.46 | 3.87 |

data to the formula

$$d' = Z_{\mathrm{N}} - aZ_{\mathrm{SN}} \qquad [1]$$

took the slope of the ROC, *a*, into account in the estimation of $d'$ from the three sets of normal deviate transforms of the HR ($Z_{\mathrm{SN}}$) and the FAR ($Z_{\mathrm{N}}$).

Two of the *S*s in the Experimental group shifted to perfect discrimination in Part 3 of the experiment. These *S*s were assigned a $d'$ score for Part 3 equal to the highest $d'$ observed in the study.

A mixed-design analysis of variance (Myers, 1966) was performed on the $d'$ values obtained for each *S*, with groups as a between-*S*s variable and Part 1 versus Part 3 as a within-*S*s variable. The main effect for groups was significant, $F\ (2, 33) = 5.07$, $p < .025$, as was the effect for Part 1 vs. Part 3, $F\ (1, 33) = 4.99$, $p < .05$. The Group × Parts interaction was also significant, $F\ (2, 33) = 5.51$, $p < .01$. Sheffé contrasts were performed on Part 1 and Part 3 means for each group with $\alpha = .05$. The contrast was significant for the conditioning group, but the contrasts for each of the Control groups failed to reach significance.

These results indicate that the classical conditioning procedures resulted in a significant increase in perceptual sensitivity. The Random-Shock Control *S*s demonstrated that this increase was not due to the arousal associated with the noxious shock. The No-Shock Control *S*s showed that the effect observed was not due to practice.

This study supports the hypothesis that classical conditioning procedures can modify the perception of stimuli. Feather and Wells (1967) demonstrated that the perceived intensity of a shock (UCS) could be distorted by the pairing of a tone (CS) with the shock. In the present study, the sensitivity of *S*s to the CS was enhanced by the pairing of the CS with a UCS, thus suggesting that the perceived intensity of the CS was increased. The outcome of this experiment concurs with the Feather and Wells observation that 5 sec. is the optimal ISI for perceptual modification.

Measures of response bias were also obtained in this study. Ingham (1970) has introduced an index of response bias, here termed $C_k$, which estimates bias in normal deviate units. The index,

$$C_k = Z_{\mathrm{N}} + Z_{\mathrm{SN}}, \qquad [2]$$

tends to be more consistent between sessions than the familiar likelihood ratio estimation of bias, and Ingham has noted that the $C_k$ index approximates an orthogonal relationship of bias to $d'$ better than the likelihood ratio. The Ingham index was obtained for each of the rating categories except Certain–Blank for each *S* in both Part 1 and Part 3, excepting the two *S*s shifting to perfect discrimination in Part 3.

To evaluate response bias changes, mean shifts in $C_k$ were obtained within *S*s across the three categories. The mean change across *S*s in the Experimental group was an increase of .27, while for the Random-Shock Control and No-Shock Control groups mean decreases of $-.33$ and $-.13$ were observed, respectively. A one-way analysis of variance for unequal *n* was performed on the mean $C_k$ shift scores, and a significant difference among the means was observed, $F\ (2, 31) = 3.88$, $p < .05$. Thus, *S*s experiencing conditioning showed an increase in willingness to report the signal during Part 3 in contrast to the Control groups, who showed decreases. This change may reflect the reaction of *S*s to their augmented sensitivity to the signal. In other words, if *S*s noted that the triangle was easier to perceive during Part 3, they may well have reacted by increasing the frequency of triangle reports on all trials.

It should be noted that the mean correlation of $d'$ and $C_k$, averaged (with Fisher $Z$ transforms) across $k_4$, $k_3$, and $k_2$ categories, and all groups, was small ($r = -.096$). Thus, sensitivity and bias estimates were orthogonal.

This study is directly relevant to the literature on perceptual defense and vigilance since it demonstrates an increase in perceptual sensitivity to a threatening stimulus when the effects of response bias are taken into account. The $d'$ increase observed here is consistent with Dorfman's (1967) finding that a higher $d'$ is evident in the perception of taboo words than in the perception of neutral words. Van Egeren (1968) reported no $d'$ difference between the affective words and neutral words, in contrast to the present study and Dorfman's experiment, but the uniqueness of his observations are probably due to basic procedural differences. Specifically, Van Egeren required *S*s to do three things: (*a*) discriminate a neutral stimulus from another neutral word, (*b*) discriminate an affective stimulus from another affective word, and (*c*) discriminate an affective stimulus word from a neutral word, or vice versa. No blank trials were used. Each $d'$ score was therefore analogous to the concept of a differential threshold or a just noticeable difference. In the Dorfman study, *S*s were given blank trials as well as threatening and neutral stimuli, so that the $d'$s reported by Dorfman were analogous to absolute thresholds; i.e., each was an index of how well *S* could distinguish between the stimulus and nothing. Similarly, in the present study, blank slides were used so that each $d'$ was conceptually similar to an absolute threshold. This experiment, Dorfman's work, and Van Egeren's study suggest that absolute sensitivity to a threatening stimulus is greater than absolute sensitivity to a neutral stimulus, but that discrimination between threatening stimuli does not differ from discrimination between neutral stimuli.

## REFERENCES

Dorfman, D. D. Recognition of taboo words as a function of a priori probability. *Journal of Personality and Social Psychology*, 1967, **7**, 1–10.

Feather, B. W., & Wells, D. T. Perceptual distortion of paired stimuli. *Psychonomic Science*, 1967, **8**, 85–86.

Goldiamond, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychological Bulletin*, 1958, **55**, 373–411.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Hake, H. W., & Rodwan, A. S. Perception and recognition. In J. B. Sidowski (Ed.), *Experimental methods and theory in psychology.* New York: McGraw-Hill, 1966.

Ingham, J. G. Individual differences in signal detection. *Acta Psychologica*, 1970, **34**, 39–50.

Myers, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

Natsoulas, T. Converging operations for perceptual defense. *Psychological Bulletin*, 1965, **64**, 393–401.

Richards, B. L., & Thornton, C. L. Quantitative methods of calculating the d′ of signal detection theory. *Educational and Psychological Measurement*, 1970, **30**, 885–889.

Van Egeren, L. Repression and sensitization: Sensitivity and recognition criteria. *Journal of Experimental Research in Personality*, 1968, **3**, 1–8.

Zajonc, R. B., & Dorfman, D. D. Perception, drive and behavior theory. *Psychological Review*, 1965, **71**, 273–290.

(Received September 29, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 343–348

# EXPECTED VALUE AS A DETERMINANT OF THE DISTRIBUTION OF ATTENTION[1]

HARVEY G. SHULMAN[2] AND RONALD P. FISHER

*Ohio State University*

The ability to control the distribution of attention across two information sources was studied in two experiments. On each trial, a 50-msec. compound stimulus consisting of a simultaneous light and tone was presented. Following stimulus presentation, *S* was cued to judge either brightness or pitch. In Exp I. the cuing probabilities for the two stimulus dimensions were varied, and in Exp. II the payoffs for correct responses were varied. In both experiments, *S*s were able to distribute their attention in proportion to the expected value of attending to each stimulus dimension.

The divided-attention, or time-sharing, paradigm has been used in at least two ways in the study of human performance. First, such methods are useful for the study of attention per se, although they have been used less for this purpose than selective attention paradigms. Lindsay (1970), for example, relied heavily on divided-attention studies in developing his theoretical approach to attention and problems of capacity sharing between multiple inputs.

A second use of divided-attention paradigms has been as a method for monitoring the information-processing load caused by the performance of a given primary task in terms of the decrement observed in the simultaneous performance of a secondary task. The basic principle involved is analogous to Archimedes' principle, which provides a method for measuring the volume of a solid object in terms of the amount of water it displaces when immersed. Johnston, Greenberg, Fisher, and Martin (1970), for example, measured the difficulty of various tasks such as stimulus encoding, retention, and response generation in terms of differences in performance obtained between a manual tracking task performed simultaneously with the task of interest and the same tracking task performed alone.

One of the critical assumptions made in using time-sharing paradigms in this way is that it is possible to control *S*'s distribution of attention to the primary and secondary tasks. Through payoffs or instructions, an attempt is made to induce the set to devote only as much capacity to the secondary task as performance on the primary task will permit. Given that a person can control the distribution of attention in this manner, changes in secondary task performance will accurately reflect the amount of change in primary task demands caused by experimental manipulations. The ability to establish control over the distribution of attention, then, is critical to the use of divided-attention tasks as measurement devices. Furthermore, the extent to which one can control the distribution of attention, and the parameters governing such control, are matters of interest in their own right, yet these have not been systematically investigated.

In both of the present experiments, the time-sharing paradigm was used to study the ability to distribute attention across multiple input sources in accordance with variations in the expected value of attending to each source. The two time-shared tasks each involved absolute judgments of unidimensional stimuli. On each trial, *S* was presented with a light and a simultaneous tone and was then cued to report either the brightness of the light or the

---

[1] This research was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant AFOSR-69-1669 to the Ohio State University Research Foundation.

[2] Requests for reprints should be sent to Harvey Shulman, Human Performance Center, 404-B West 17th Ave., Columbus, Ohio 43210.

pitch of the tone. In various conditions of Exp. I, the probabilities that *S* would be cued on each modality were varied, with the payoff for each correct report held constant. In Exp. II, the cuing probabilities were held constant and the monetary payoffs for correct responses on each modality were varied. The behavior of interest in each experiment was the ability of *S* to allocate his processing capacity in proportion to the expected value of attending to each dimension.

## EXPERIMENT I

### *Method*

*Procedure.*—The nature of the stimulus event on each trial was the same in all experimental conditions. Each stimulus consisted of the simultaneous presentation of a 50-msec. light flash and a 50-msec. tone. The visual stimulus could have any one of four intensity values, the auditory any one of four frequency values. Thus 16 possible two-dimensional stimuli were possible, and were each used with equal frequency in each condition. The *S* initiated each trial by pressing a hand-held button. One second later the stimulus was presented and was followed after 1 sec. by a cue informing *S* that he was to report either the brightness of the light or the pitch of the tone. Immediate feedback was provided by *E* after each trial.

Verbal responses were assigned to each of the four visual intensity levels, and auditory frequency levels used as stimuli. The intensity levels of the visual stimuli and their labels were 24 ftc. ("one"), 16 ftc. ("two"), 11 ftc. ("three"), and 7 ftc ("four"). The four tones were equated for loudness at 72 db. and had frequencies of 1,000, 960, 920, and 880 Hz., assigned the verbal labels of "A," "B," "C," and "D," respectively.

In the initial five sessions of the experiment, each *S* was given practice at the experimental task, in order to provide adequate familiarity with the verbal labels and the time-sharing paradigm. Each then served in five sessions for Exp. I, followed by the five sessions of Exp. II. Each session consisted of 192 trials in a single experimental condition and lasted approximately 40 min.

*Design.*—In Exp. I, each *S* served in five experimental conditions differing from each other in terms of the proportions of trials on which the tested dimension was drawn from each modality. In two control conditions, the same dimension was tested after each stimulus presentation. Hence, in one condition, *S* was always asked to judge brightness, never pitch, and in another always judged pitch, never brightness. These conditions permit selective attention to the cued dimension and therefore provide base-line levels of performance for each dimension. In a third condition, each dimension was tested on 50% of the stimulus presentations, with the order of tested dimensions randomized within each session. This condition therefore required *S* to distribute his attention equally across the two dimensions in each stimulus. In the fourth (and fifth) condition, the proportion of trials on which the visual (auditory) dimension was tested was .75, and the auditory (visual) dimension was tested with a probability of .25. In all conditions of Exp. I, a correct response on each trial was worth .3¢ in bonus earnings. The *Ss* were informed of the cuing probabilities assigned each dimension at the beginning of each trial block in a condition.

Each session consisted of 192 trials in a single experimental condition. The assignment of conditions to sessions was counterbalanced by assigning two *Ss* to each row of a 5 × 5 Latin square. Each of the 16 possible two-dimensional stimuli occurred 12 times in each session, and no one brightness level or frequency was tested more often than any other. The presentation order of stimuli and the order of tested dimensions were randomized separately for each session.

*Apparatus.*—The *S* was seated at a small table approximately 1 m. away from an opaque Plexiglas panel whose surface was covered with black construction paper. In the middle of this panel was an opening 2.54 cm. on which the visual stimulus was back projected. Auditory stimuli were presented binaurally over Wilson sound barrier headphones. At the base of the Plexiglas panel were two indicator lights labeled frequency and brightness. These were used to cue *S* as to which dimension he was to report on after each stimulus presentation. Visual stimuli were presented with a Kodak Carousel random access projector using slides made from Wratten neutral density gelatin filters (Kodak, No. 96) having densities of .40, .60, .80, and 1.0. Auditory stimuli were provided by a set of four tone generators, calibrated for fequency prior to each experimental session. The timing of events on each trial was controlled by a Hunter (Model 1516) six-channel timer, which controlled a shutter on the projector and also gated the onset and offset of the auditory stimulus.

*Subjects.*—Five male and five female undergraduate students volunteered as *Ss* in response to an advertisement in the Ohio State University newspaper. Each was paid a base rate of $1.25 per session in addition to the bonus pay described earlier.

### *Results*

For each *S*, the proportion of correct judgments achieved on each modality in each condition was calculated. Table 1 shows these data averaged over *Ss*. Note that Table 1 (as well as Table 3) does not reflect that feature of the experimental design which required that each cuing probability in one modality be paired with

TABLE 1

Proportion Correct as a Function of Cuing Probability

| Dimension | Cuing probability |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|
|  | .25 |  | .50 |  | .75 |  | 1.00 |  |
|  | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ |
| Brightness | .60 | .07 | .68 | .06 | .70 | .10 | .72 | .09 |
| Pitch | .75 | .11 | .76 | .09 | .77 | .09 | .83 | .11 |
| $\bar{X}$ | .68 |  | .72 |  | .74 |  | .78 |  |

TABLE 2

Transmitted Information Scores in Bits for Each Condition of Experiment I

| Dimension | Cuing probability |  |  |  |
|---|---|---|---|---|
|  | $1.0^{a,b}/.0^{c}$ | $.75^{a}/.25^{b}$ | $.50^{a}/.50^{b}$ | $.25^{a}/.75^{b}$ |
| Brightness | 1.02 | .98 | .90 | .77 |
| Pitch | 1.30 | 1.14 | 1.14 | 1.18 |
| Total | 2.32 | 2.12 | 2.04 | 1.95 |

[a] Cuing probability for brightness.
[b] Cuing probability for pitch.
[c] Data in this column are from the two control conditions which would properly be $1.0^{a}/.0^{b}$ and $.0^{a}/1.0^{b}$.

the complementary cuing probability in the other modality. For example, in the block of trials where brightness was the cued dimension 75% of the time, pitch was cued 25% of the time.

An analysis of variance was performed on the proportion of correct responses, which has as factors modality, cuing probability (either .25, .50, .75, or 1.00), and *S*s. The main effect of modality was significant at $p < .05$, with $F\ (1, 9) = 7.79$. The main effect of cuing probability was also significant, $F\ (3, 27) = 9.25$, $p < .001$, as was the Modality × Cuing Probability interaction, $F\ (3, 27) = 5.10$, $p < .01$.

The main effect of modality reflects the overall superiority of performance on the auditory judgment task. The Modality × Cuing Probability interaction is due to a difference in the shapes of the functions relating cuing probability to performance in the two modalities. Table 1 also shows the main effect of cuing probability for both modalities combined. The proportion of correct judgments increased monotonically with cuing probability. Presumably this effect reflects the allocation of processing capacity to stimulus dimensions in proportion to *S*'s changing expectation as to which dimension he will be tested on.

Performance was also scored in terms of $d'$ scores and transmitted information since Taylor, Lindsay, and Forbes (1967) and Tulving and Lindsay (1967) have used these measures in similar contexts. The use of each of these measures resulted in no difference in the pattern of results, and significance tests obtained using proportion correct as the dependent variable. Transmitted information ($H_t$) scores are nonetheless reported in Table 2, which presents the averages taken after computing each *S*'s $H_t$ score in each condition. Unlike proportion correct, $H_t$ has the desirable property of permitting one to sum $H_t$ scores over conditions, providing a potential means of determining whether the total amount of discrimination ability remains constant across conditions differing from each other in the manner in which attention was distributed. The third row of Table 2 presents total $H_t$ for each cuing condition. These data indicate a progressive improvement in total performance as more weight is assigned to the visual dimension. An analysis of variance done on total $H_t$ with *S*s and conditions as factors revealed a significant main effect of conditions, $F\ (3, 27) = 7.06$, $p < .01$. The control condition resulted in a higher total $H_t$ score than in any of the conditions requiring divided attention, a result which might be taken to indicate a decrease in the efficiency of performance as one changes from selective to distributed attention. However, this inference is best regarded as tentative since the result on which it is based was not replicated in Exp. II.

Since the payoff for a correct response was held constant across levels of cuing probability, the expected value of attending to each dimension of the compound stimulus covaried with cuing probability. The hypothesis that expected value can provide a criterion for capacity allocation is an attractive one and may be converged upon by holding cuing probability contant ands varying the payoffs assigned to the two stimulus dimensions. This manipulation was performed in Exp. II.

## EXPERIMENT II

### *Method*

*Design and procedure.*—The procedure was identical with Exp. I, as were the details of the experimental design. The difference between Exp. I and II was solely in terms of the method used to manipulate the expected value of attending to the two stimulus dimensions. In all conditions, the cuing probability for each dimension was .50. The five conditions differed from each other in terms of the payoff values assigned to a correct judgment on each dimension. In one condition, the visual dimension was assigned a payoff value of .6¢ per trial while the auditory had a value of 0¢ per trial. In a second condition, these payoffs were reversed. The third condition paid equal amounts (.3¢) for correct responses on either dimension, and in the fourth and fifth condition values of .45¢ and .15¢ were assigned to the two dimensions, with one condition having the higher payoff for the visual dimension, the other for the auditory.

*Subjects.*—The same 10 *S*s who served in Exp. I also participated in Exp. II. Their pay was again $1.25 per session plus whatever performance bonus they earned.

### *Results*

The data were subjected to an analysis of variance similar to that used in Exp. I. The one difference was that in Exp. II data were available from both modalities in all five payoff conditions, whereas in Exp. I the two extreme conditions provided no data on the dimension whose cuing probability was zero. In Exp. II, the main effects of both modality and payoff level were significant, $F\ (1, 9) = 8.60, p < .025$, and, $F\ (4, 36) = 8.24, p < .001$, respectively. The interaction between modality and payoff level did not approach significance, $F\ (4, 36) = 1.95$.

Table 3 shows the relationship between the payoff value assigned to a stimulus dimension and performance on that dimension, averaged over modalities. The relationship is similar to that of Exp. I, although less well described by a linear function. Table 3 also gives the proportion of correct judgments made in each payoff condition for each modality. In each modality, there is a monotonic increase in performance with increasing payoff values with the exception that performance under the two highest payoffs was nearly equal. This may well reflect a ceiling on performance set by the discriminability of the four points chosen on each dimension for use as stimuli.

Table 4 presents $H_t$ scores for Exp. II. As in Exp. I, the effect of conditions on total $H_t$ was significant by analysis of variance, $F\ (4, 36) = 3.80, p < .05$. There again appears to be an increase in total $H_t$ as the weighting of brightness increases. However, in Exp. II, unlike Exp. I, the total $H_t$ scores in the two control conditions are lower than in the conditions requiring distributed attention. We are unable to suggest an explanation for this discrepancy.

Finally, the data from conditions having the same expected value for each correct judgment in Exp. I and II were combined. In Fig. 1, the proportion of correct judgments in each of these conditions is plotted for brightness and pitch, as a function of expected value. These data illustrate two generalities which cover both experiments. First, pitch judgments were made

TABLE 3

PROPORTION CORRECT AS A FUNCTION OF PAYOFF VALUE

| Dimension | Payoff value in cents | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | | .15 | | .30 | | .45 | | .60 | |
| | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ |
| Brightness | .56 | .16 | .68 | .09 | .71 | .09 | .77 | .05 | .76 | .08 |
| Pitch | .69 | .14 | .75 | .12 | .77 | .08 | .81 | .09 | .80 | .06 |
| $\bar{X}$ | .63 | | .72 | | .74 | | .79 | | .78 | |

TABLE 4

Transmitted Information Scores in Bits for Each Condition of Experiment II

| Dimension | Payoff value in cents | | | | |
|---|---|---|---|---|---|
| | .60[a]/0[b] | .45[a]/.15[b] | .30[a]/.30[b] | .15[a]/.45[b] | 0[a]/.60[b] |
| Brightness | 1.02 | 1.12 | 1.00 | .91 | .69 |
| Pitch | .98 | 1.15 | 1.16 | 1.31 | 1.26 |
| Total | 2.00 | 2.27 | 2.16 | 2.22 | 1.95 |

[a] Payoff value in cents for each correct brightness judgment.
[b] Payoff value in cents for each correct pitch judgment.

with greater accuracy than brightness judgments, probably because the visual stimuli were less discriminable from each other than were the auditory stimuli. This interpretation is based upon the superiority of pitch judgments over brightness judgments in the control conditions of both experiments, which did not require divided attention. One might ask for a control condition in which only a single stimulus dimension was presented in order to support the inference that one dimension was easier than the other. This type of control condition was not included in the present experiments because Tulving and Lindsay (1967) found no difference between conditions in which visual and auditory stimuli are simultaneously presented and only one is attended to and judged and conditions in which a single stimulus is presented and judged. The implication is that the present control conditions index the upper level on performance possible with the present stimuli in the same way that a single stimulus control condition would.

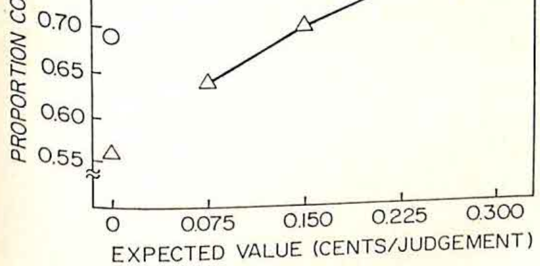


Fig. 1. Proportion correct judgments as a function of expected value in Exp. I and II combined. (The points plotted for an expected value of zero are based on Exp. II only.)

The second point to be made with reference to Fig. 1 is that the slope of the function relating performance to expected value is greater for the visual dimension than for the auditory. It was this fact which produced the significant Modality × Cuing Probability interaction in Exp. I. If one regards the slope of the functions in Fig. 1 as an index of the effectiveness of reallocating processing capacity to each dimension, then the inference to be drawn is that capacity reallocation has a larger effect on the visual than on the auditory dimension.

## Discussion

The results of both experiments support the conclusion that the distribution of attention between two sources of information may be controlled by *S*, in an apparently rational manner. This finding, then, provides support for the rationale adopted whenever divided attention paradigms are used to measure the cognitive load imposed by a primary task in terms of performance variations on a concurrent secondary task.

There are, however, a number of important questions remaining to be answered with respect to the distribution of attention. Assume first that a person is to some extent able to assign differential weights to potential sources of information, and then distribute his attention according to these weights. The present results indicate that the expected value of processing a source of information is one way in which such weights can be determined. Undoubtedly there are other such determinants, and the discovery of their functional relationships with attention is an important problem. In the literature on divided attention, for example, there is some evidence that *S*'s perception of the relative difficulty of the two tasks influences the manner in which he distributes attention. Mowbray (1953) found that the easier of two tasks consistently underwent the greater time-sharing decrement in comparison with single-task control performance. Klemmer (1958) equated the difficulty of auditory and visual tasks which were then time shared. Klemmer found, as did Mowbray, that the auditory task underwent the greater decrement when time shared, suggesting that under some circumstances *S*s may assign a greater weight to the visual modality.

The present results indicate that the pitch judgment task in the present experiment was less difficult than the brightness judgment task. Furthermore, if one compares performance in the control conditions of Exp. I and II to performance in any of the divided-attention conditions, a greater decrement is observed for pitch judgments than brightness. Although these particular findings are consistent with Mowbray (1953) and Klemmer (1958), there is need for caution in accepting the generality of the conclusion that the auditory modality will always undergo a greater time-sharing decrement than the visual modality. This follows from the possibility that the control conditions in Exp. I and II for brightness judgments may be constrained by a ceiling on performance imposed by the discriminability of the brightness values chosen as stimuli. If one ignores these control conditions and attends instead to the slope of the functions in Fig. 1, the visual dimension appears to show the greater time-sharing effect. The relationship between task difficulty, task modality, and time sharing, then, requires further exploration.

There is also evidence, in the literature on selective attention, that certain stimuli are assigned permanently high weights, and the distribution of attention to such stimuli appears to be mandatory rather than under *S*'s control. Moray (1959) found that *S*s asked to attend to one verbal message and ignore another presented simultaneously were able to do so quite effectively, unless the ignored message was prefixed by *S*'s own name. The extent to which the distribution of attention is under voluntary control and the situational determinants of voluntary control are both relatively unexplored areas in which research is desirable.

Finally, it is worth noting that in recent years much of the research done on attention has focused on two related issues. The first is whether the selective nature of attention is best represented by models based on analogies with a discrete switching mechanism (Broadbent, 1958) or a continuous attenuating mechanism (Treisman, 1969). The second is whether this selective mechanism is perceptual or at the level of response selection (Deutsch & Deutsch, 1963). None of the positions taken on these issues has been particularly susceptible to disproof by experimentation, and an unfortunate consequence has been that other problem areas having to do with attention have been largely ignored. The present research implies that attention may be rationally distributed, a finding that falls outside the scope of current models of attention whether they are based on filtering or attenuating, and whether they are selective at the level of sensory analysis or response organization. However, no theory of attention can be considered adequate in any sense until it incorporates a set of statements about the determinants of capacity allocation, as well as the nature of the selection process that takes place when one input is processed and another ignored.

## REFERENCES

BROADBENT, D. D. *Perception and communication.* London: Pergamon Press, 1958.

DEUTSCH, J., & DEUTSCH, D. Attention: Some theoretical considerations. *Psychological Review,* 1963, **70**, 80–90.

JOHNSTON, W. A., GREENBERG, S. N., FISHER, R. P., & MARTIN, D. W. Divided attention: A vehicle for monitoring memory processes. *Journal of Experimental Psychology*, 1970, **83**, 164–171.

KLEMMER, E. T. Time-sharing between frequency coded audio and visual channels. *Journal of Experimental Psychology*, 1958, **55**, 229–235.

LINDSAY, P. H. Multichannel processing in perception. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis.* New York: Appleton-Century-Crofts, 1970.

MORAY, N. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology,* 1959, **9**, 56–60.

MOWBRAY, G. Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty. *Journal of Experimental Psychology*, 1953, **46**, 365–372.

TAYLOR, M. M., LINDSAY, P. H., & FORBES, S. M. Quantification of shared capacity processing in auditory and visual discrimination. In A. F. Sanders (Ed.), *Attention and performance.* Amsterdam: North Holland Publishing Co., 1967.

TREISMAN, M. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–299.

TULVING, E., & LINDSAY, P. H. Identification of simultaneously presented simple visual and auditory stimuli. In A. F. Sanders (Ed.), *Attention and performance.* Amsterdam: North Holland Publishing Co., 1967.

(Received September 30, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 349–353

# STIMULUS GENERALIZATION VARIABLES IN ANAGRAM-PROBLEM SOLVING

ERNEST H. LEMAY[1]

*Plattsburgh State University College*

Two factorial experiments were run which varied the position and serial order of the letters in the anagrams. Letter position was varied by having 0, 1, or 2 letters in their correct position in Exp. I and 0 or 2 in Exp. II. Serial order was varied by having 0 or 2 letters in proper serial order for both Exp. I and II. Experiment I used a between-*S*s design, while Exp. II used a within-*S*s design. Analysis of the data revealed that letter position and serial order were effective variables in both Exp. I and II. An analysis of covariance on the Exp. I data showed that serial order and letter position were closely related. It was concluded that both letter position and serial orders are variables which affect stimulus generalization.

Anagram-problem solving has been viewed as occurring in two phases (Hunter, 1959; Johnson, 1966). First, *S* rearranges the letters of the anagram and second, this rearrangement is compared to some word models which *S* has stored in his memory system. LeMay (1972) has suggested that this latter phase involves stimulus generalization. Not all attempted rearrangements of the letters of the anagram will exactly duplicate the actual solution word. It is apparent that *S*s can solve anagrams even when the rearranged letters only bear a similarity to the actual solution word. Thus, a problem in stimulus generalization occurs.

Two variables stand out in studying stimulus generalization. In constructing an anagram, some of the solution-word letters can appear in their correct position, or they can be in their correct serial order. For example, if the solution word is ABCDE, then $\underline{\text{A}}$CBED would be an example of the former case and is referred to in this study as the letter-position variable. (The letter underlined is in its correct position.) The anagram BA$\overline{\text{DE}}$C would be appropriate for the latter category and is referred to as the serial-order variable. (The letters overlined are in the correct serial order.) It is possible to arrange these two variables in a factorial design. There are 120 possible combinations of five letters. Of these, there are 14 anagrams in which neither letter position nor serial order are maintained. The condition which contains anagrams constructed in this way will be referred to as 0–0, where the first number refers to the letter-position variable and the second number to the serial-order variable. There are 28 anagrams which fall in the 1–0 category; i.e., only one letter is in its correct position, and there are no serial orders. The 2—0 category contains 10 arrangements; 0–2, 22; 1–2, 10; and there are 6 in the 2–2 category. Other combinations are possible but either do not exist in sufficient numbers or cannot be incorporated in a factorial design. Thus, the present study factorially combines three levels of letter position (0, 1, and 2) with two levels of serial order (0 and 2). Five arrangements from each of the above categories were used in each condition of Exp. I and IV in Exp. II. A third variable, i.e., solution-word-bigram frequency total, was included, because it is a solution-word characteristic which might bear some importance to stimulus generalization. It was hypothesized that in those conditions where serial order was present, i.e., 0–2, 1–2, and 2–2, high-bigram frequency totals would facilitate *S*'s "perception" of the correct letter pairs; thus, an interaction should be present.

Several studies have been concerned with

---

[1] The author wishes to thank Susan Champagne for her assistance in running *S*s and analyzing the data.

Requests for reprints should be sent to Ernest H. LeMay, Department of Psychology, Plattsburgh State University College of Arts and Science, Plattsburgh, New York 12901.

the arrangements of the anagram letters in problem solving (e.g., Dominowski, 1966, 1968; Hunter, 1959; Mayzner & Tresselt, 1958); however with the exception of the Dominowski (1968) study, none of these accurately delimited serial-order and letter-position variables.

Dominowski (1968) reported an experiment which was somewhat similar to the present study. In his experiment he provided *S*s with letter-sequence information; i.e., *S*s were informed of either the correct position of a single letter, the correct serial order of two or three letters, or of the serial order and correct position of two letters. In the sense of the information provided, he used the following groups: 0–0 for the control, i.e., neither letter-position nor serial-order information was given to *S*s, 1—0, 0—2, 2—2, and 0—3. It was found that serial order was a more potent variable than letter position.

The present study differed from Dominowski's (1968) in that (*a*) *S*s were not given specific letter-sequence information, and (*b*) a factorial design was used combining two levels of serial order with three levels of letter position. In the Dominowski (1968) study, *S*s were directed to the relevant information, while in the present study, they had to be able to pick out the correct letters without aid from the instructions. The question to be answered is: "When does an anagram look like the solution word?"

## EXPERIMENT I

### *Method*

*Subjects and design.*—The 240 *S*s used in this experiment were volunteer Plattsburgh State University College freshmen taking a general psychology course. There were 12 conditions in the experiment corresponding to a 2 × 3 × 2 factorial design combining 2 levels of solution-word-bigram frequency total (high vs. low), 3 levels of letter position (0, 1, and 2), and 2 levels of serial order (0 and 2). The *S*s were run in large groups, and a randomized block design was used in assigning *S*s to the condition.

*Materials.*—Twenty high-frequency (AA) five-letter words were selected from the Thorndike and Lorge word book. All had single solutions according to the Olson and Schwartz (1967) list. Ten of the solution words had high bigram totals ($\bar{X}$ = 5,910.6; range = 4,640 to 7,234) while the other 10 words had low bigram totals ($\bar{X}$ = 2,553.3, range = 1,283 to 3,041).

There were 120 anagrams constructed from these words, 10 for each combination of letter-position, serial-order, and bigram-total variable. Within an anagram condition, five arrangements of the letters were used. Thus, within a particular condition, an arrangement occurred twice. The words were paired with the arrangements on a random basis. Mean anagram–bigram frequency totals varied only slightly across conditions, from 2,627.7 to 2,992.0.

The anagram arrangements were further quantified for purposes of analyses by computing the number of moves necessary to reach solution and the number of inversions in each anagram. When using the number of moves to solution, the distance moved is disregarded (Dominowski, 1966). By using the number of inversions, the distance is taken into consideration. Given the series 12345, an inversion occurs whenever a larger number proceeds a smaller number (Hohn, 1964, p. 43). For example, the series 24153 contains four inversions, namely 2 and 4 occur before 1, and 4 and 5 occur before 3.

The anagrams were printed on sheets. Each sheet consisting of a particular condition of 10 anagrams. Space was left for writing the solutions.

*Procedures.*—The experiment was run in a group setting. The *S*s were read the standard instructions for anagram–problem solving and were shown an example of an anagram. A time limit of 8 min. was allowed for solving the problems. Each *S* was exposed to only one condition.

### *Results and Discussion*

The results of this study are presented graphically in Fig. 1. A 2 × 3 × 2 analy-

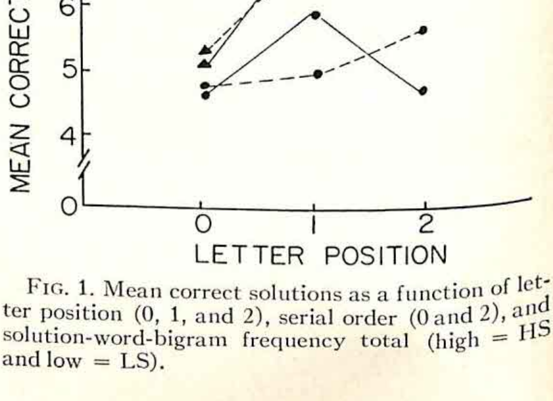


FIG. 1. Mean correct solutions as a function of letter position (0, 1, and 2), serial order (0 and 2), and solution-word-bigram frequency total (high = HS and low = LS).

sis of variance was carried out on the data of Exp. I. This analysis showed that the solution-word-bigram frequency total was not an effective variable, producing no main effects or interactions. Thus, this variable will not be discussed further. The letter-position variable was effective, $F\ (2,\ 228) = 11.51$, $p < .01$, with performance improving as the number of correct letter positions increased. As serial order increased, so did the number of solutions, $F\ (1,\ 228) = 25.66$, $p < .01$. The only interaction was Letter Position × Serial Order, $F\ (2,\ 228) = 5.72$, $p < .01$, such that there was a sharper rise as the letter-position variable increased for the higher serial-order level than for the lower level of serial order. A trend analysis was carried out on these data and revealed a linear letter-position trend, $E\ (1,\ 228) = 21.45$, $p < .01$, which confirmed the finding concerning this variable mentioned above. A linear trend for the Letter Position × Serial Order interaction was found, $F\ (1,\ 228) = 11.29$, $p < .01$. This shows that the two slopes, one for the (0)-serial-order condition and that for the (2) condition are both linear but increase differentially.

Both letter position and serial order affected the solution of the anagrams; however, because it is difficult to manipulate these variables without also affecting the number of letter moves necessary to obtain a solution. The data were examined to determine (*a*) if the distance moved is important, and (*b*) the extent which the number of letter moves affected the solutions.

A correlation analysis showed that letter moves were correlated with solutions, $r\ (29) = -.54$, $p < .01$; while the correlation between the number of inversions and solutions was not significant, $r\ (29) = -.31$, $p > .05$, although in the right direction.

Because of this, an analysis of covariance was run on these data with the number of letter moves used as the covariate. To accomplish this, the number of solutions per word, per condition was calculated.

The analysis of covariance produced only one significant finding and that was for the letter–position variable, $F\ (2,\ 17) = 7.11$, $\mathrm{p} < .01$. Thus, it would appear that serial order and the number of letter moves are measuring much of the same thing. On the other hand, the letter-position variable cannot be accounted for on the basis of letter moves.

One further analysis was carried out on these data. In Cond. 2—2, it was possible for letter position and serial order to apply to the same letters, e.g., $\mathrm{E}\underline{\overline{\mathrm{BC}}}\mathrm{AD}$. It was also possible to keep these two variables distinct, e.g., $\underline{\mathrm{A}}\overline{\mathrm{CD}}\mathrm{B}\underline{\mathrm{E}}$. It is of interest to know which type of arrangement facilitated solutions most. The proportion correct in each category was computed, and an arcsine transformation was carried out on these proportions.

A one-way analysis of variance for repeated measures was run on these data. The results of this analysis showed that the mean number of solution for the $\underline{\mathrm{A}}\overline{\mathrm{CD}}\mathrm{B}\underline{\mathrm{E}}$ ($\bar{X} = 4.62$ adjusted mean based on five anagrams in each category) paradigm was greater, $F\ (1,\ 39) = 17.35$, $p < .01$, than that for the $\mathrm{E}\underline{\overline{\mathrm{BC}}}\mathrm{AD}$ paradigm ($\bar{X} = 4.02$ adjusted mean).

Finally, when the students were questioned at the end of the experiment regarding whether they had been aware of anything peculiar about the anagram, five *S*s in Cond. 2—2 and three in Cond. 1—2 stated that they had noticed that some of the letters were in their correct position or order. It was suspected that *S*s might have developed a set or strategy in solving the anagrams, to wit, "look for combinations of letters." Therefore, Exp. II was run to eliminate this possibility.

## EXPERIMENT II

### *Method*

*Subjects and design.*—The *S*s were 40 Plattsburgh State University College freshmen taking a general psychology course, who volunteered for the experiment. These *S*s were run in a 2 × 2 factorial design which varied 2 levels of letter position (0 and 2) with 2 levels of serial order (0 and 2), which resulted in four conditions, 0–0, 2–0, 0–2, and 2–2. A repeated-measure design was used where all the *S*s experienced all of the conditions.

*Materials and procedures.*—The words used in Exp. II were the same as those used in Exp. I. The *S*s were required to solve 16 anagrams, 4 from each of the above conditions. Two lists of anagrams were prepared and differed in that the same word was not allowed to appear in the same condition on both lists. Thus, the word-anagram pairs were different on the two lists. Two random orders of each of these lists were prepared, resulting in four sheets of 16 anagrams each. Ten *S*s were run on each sheet and a randomized-block design was used in assigning *S*s to the sheets. Experiment II did not use the same word-anagram pairs as used in Exp. I. A total of 10 min. was allowed *S*s to solve the 16 anagrams; otherwise, the procedures were the same as in Exp. I.

### *Results and Discussion*

The results of the experiment showed the means and standard deviations for the conditions to be as follows: for Cond. 0–0, $\bar{X} = 1.75$, $s = 1.19$; 2–0, $\bar{X} = 2.17$, $s = 1.37$; 0–2, $\bar{X} = 2.45$, $s = 1.43$; and 2–2, $\bar{X} = 2.75$, $s = 1.01$. The raw scores were transformed to angles, and a two-way analysis of variance was carried out on these data. Letter position was an effective variable, $F$ (1, 39) = 6.09, $p < .05$, as well as serial order, $F$ (1, 39) = 23.04, $p < .01$. The interaction between these two variables was not significant however, $F$ (1, 39) = .28, $p > .05$. The major difference being due to serial order, $\bar{X}$ difference = .69, while that for letter position was $\bar{X}$ difference = .36. The former being almost twice as large.

Thus with the exception of the interaction between the two major variables, Exp. II presents findings similar to those of Exp. I. It would appear that the interaction in Exp. I could be due to a set developed by *S*s through self-instruction, which facilitated primarily the 1–2 and 2–2 conditions.

The *S*s of this study were also questioned regarding the nature of the anagrams. Several *S*s reported that some of the anagrams were easy because "they looked like the solution word."

## GENERAL DISCUSSION

Both letter moves and serial order were effective variables in this study. These results are in general consistent with the findings of Dominowski (1968), who showed that better performance resulted as the anagram cues provided more information about the solution word. Although Dominowski (1968) did not find single letter-position cues to be effective, there were strong indications in the present study that letter position is an effective variable but not necessarily one which produces large differences.

From these findings, it is clear that viewing this problem as an issue in stimulus generalization should prove to be a fruitful approach. As the anagram becomes more like the solution word in either serial order or letter position, a solution is more likely. This implies that *S*s react to the correct letters in solving anagram problems. Although an analysis of covariance demonstrated that the effects of serial order were closely related to the number-of-letter-moves variable, there are problems with this latter concept which necessitates that it not be viewed as the crucial dimension. It is quite possible that these two variables refer to the same thing; however, the notion of serial order is more plausible for the following reasons. The *S*s do not know which letters are to be moved. To the extent that this is true, then, the letter-moves variable implies that *S*s will randomly permute the arrangements until a solution is found, but this also seems unlikely. Actually, the number of letter moves really measures the number of letters which are incorrectly placed which should be inversely related to the number of letters which are in correct serial order.

The present notion concerning stimulus generalization considers the wrong letters as being irrelevant in producing the solution word. The *S* arrives at the solution word, because it is suggested by the correct letters. The wrong letters only serve to substantiate the solution. The *S*s should do reasonably well in solving anagrams where the irrelevant letters are missing, e.g., GRP____ = group, including OU eliminates GROPE as a solution.

Support for this notion is gleaned from the fact that inversions do not add any information which increases the accuracy of prediction over letter moves. Thus, if letter moves were important, it would seem reasonable that the distance moved should also be important. If a letter is not in its correct position or serial order, it does not matter how far it is which underscores the notion that it is irrelevant. Another point is that in Cond. 2–2, *S*s found $\underline{A}\overline{CD}\underline{B}E$ type anagrams easier to solve than $E\underline{\overline{BC}}AD$ types. There are more correct letters in the former category than in the latter. Lastly, Kaplan and Carvellas (1968) found that increases in anagram length did not produce increases in solution times in direct relation to what would be expected by the increase in the total number of permutations.

They accounted for this in part on the basis of familiar affixes in longer words which were perceived by *S*s and, hence, enhance solutions. These findings argue against the letter moves notion on the one hand, because there should be a greater number of letter moves as the anagram length increases and, on the other hand, to the extent that *S*s do recognize affixes, argues for stimulus generalization.

## REFERENCES

Dominowski, R. L. Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 107–111.

Dominowski, R. L. Anagram solving as a function of letter-sequence information. *Journal of Experimental Psychology*, 1968, Vol. **76**, 78–83.

Hohn, F. E. *Elementary matrix algebra*. New York: Macmillan, 1964.

Hunter, I. M. L. The solving of five-letter anagram problems. *British Journal of Psychology*, 1959, **50**, 193–206.

Johnson, D. M. Solution of anagrams. *Psychological Bulletin*, 1966, **66**, 371–384.

Kaplan, I. T., & Carvellas, T. Effect of word length on an anagram solution time. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 201–206.

LeMay, E. H. Anagram solutions as a function of task variables and solution-word models. *Journal of Experimental Psychology*, 1972, **92**, 65–68.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology*, 1958, **56**, 376–379.

Olson, R., & Schwartz, R. Single and multiple solution five-letter words. *Psychonomic Monograph Supplements*, 1967, **2** (8, Whole No. 24).

(Received October 2, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 354–356

# A CULTURE-FREE LEARNING TASK

B. R. BUGELSKI[1] AND SANDRA LATTANZIO

*State University of New York at Buffalo*

College student *S*s learned eight nonsense syllables as a serial list and eight, two-arm positions as another serial list. The verbal learning was faster than the "motor" learning, but both types of materials showed the typical serial learning effect. The semaphore arm-movement activity is recommended as a culture-free learning task and as a suitable task for nonverbal human *S*s.

The introduction of nonsense syllables by Ebbinghaus was originally regarded as an answer to a felt need for experience-free learning materials. Subsequent experience resulted in the discovery of wide-ranging association values with large individual differences among *S*s. Currently the vogue is to dispense with nonsense materials altogether in favor of ordinary words evaluated by frequency counts, $m$ values, and most recently by imagery ($i$) values (Paivio, Yuille, & Madigan, 1968).

The acceptance of meaningful material for studies of learning and memory is fruitful, but there may still be interest in and use for a nonmeaningful activity where learning can be studied especially with *S*s who are nonverbal for any reason or where verbal experience is impoverished or in some language unfamiliar to the researcher.

A possible source of material for use with such *S*s or where verbalization is not an issue is the use of arm movements such as are used in a semaphore code. Such a code is unfamiliar to most people, while anyone capable of moving his arms about can be shown an arm position and instructed verbally or by suitable pantomime to imitate the indicated arm positions. The task, while superficially a perceptual-motor affair, does not strain either perceptual or motor capacities of any unhandicapped people. It is not a case of motor learning as examplified by pursuit-tracking or even runway or maze activity; rather, it is a task clearly comparable with verbal learning of lists or paired associates. An early study by Sheffield (1946) involved a paired-associate task in which verbal and motor movements were associated. Sheffield's movements, however, called for a variety of activities which appear to lack the comparability of the present proposal.

In a preliminary exploration of this type of learning task, *S*s were asked to learn eight arm positions in serial order. The learning of the motor movements was compared with the learning of eight nonsense syllables to observe any differences. The experimental program is described below.

## METHOD

*Subjects.*—Twenty-five college juniors served as *S*s for this study.

*Procedure.*—As each *S* entered the experimental room he was assigned to Group A or B. Group A *S*s learned nonsense syllables first and motor movements second—Group B *S*s learned the materials in a reversed sequence. Both nonsense materials and motor tasks were displayed on a screen for 3-sec. intervals with 10 sec. between trials. A Kodak Carousel projector was used to present materials. Both verbal and pictorial items were randomized for each *S*. The interitem time was automatically controlled to the time necessary to change slides. The motor task was displayed as a series of mannequins holding flags (see Fig. 1). Both kinds of learning followed anticipation instructions. The criterion of learning was one successful trial. As soon as one task was completed, the other was started with about a 1-min. rest interval.

## RESULTS

The motor task proved more than three times as difficult as the verbal task. The data appear in Table 1. The relative difficulty cannot be evaluated with these data because the possible association values of arm positions have not been appraised. The nonsense syllables used were relatively

---

[1] Requests for reprints should be sent to B. R. Bugelski, Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14226.

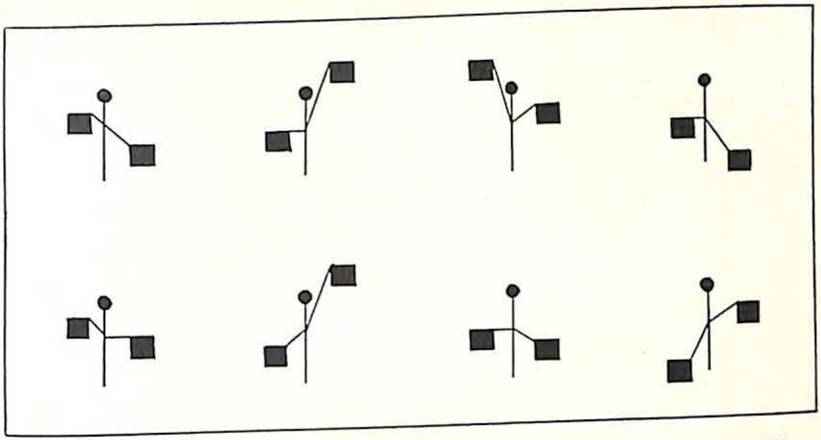


Fig. 1. The mannequin display. Shown are 8 of the 25 possible combinations using only the arm positions illustrated. (In the learning task, *S* faces the screen and mimics the positions. The *E* observes *S* from the rear and compares *S*'s arm positions with those on the screen.)

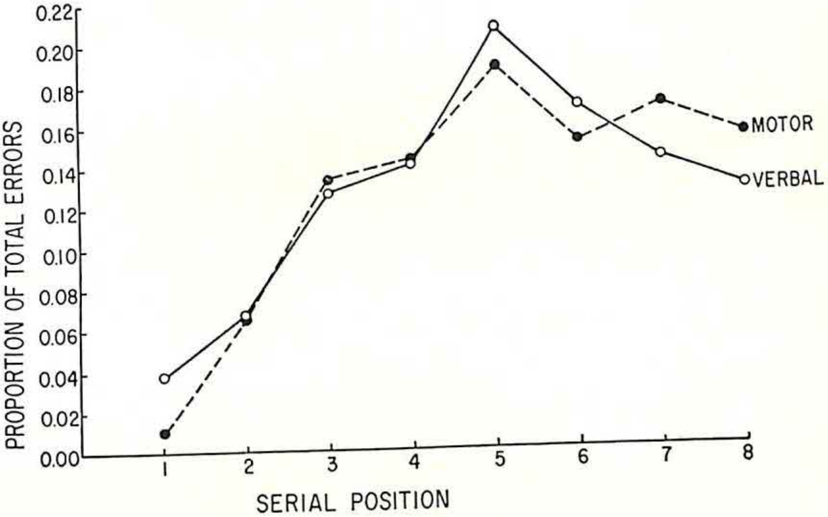


Fig. 2. The serial position effect for motor and nonsense syllable materials.

easy (about 38% association value). Some of the *S*s admitted in interrogation that they tried to verbalize the arm positions; others denied any such mediational attempts. Comparison of verbalizers and nonverbalizers showed no differences in scores or serial position effects.

When the errors are plotted by serial position, the motor learning and verbal learning appear to follow the same order with the typical difficulty in the middle of the task. See Fig. 2.

## Discussion

The fact that the serial order curves are virtually the same for the two kinds of learning

TABLE 1

Mean Trials to Criterion and Mean Error Scores for Verbal and Motor Learning Tasks

| Task | Mean trials | *SD* | Range | Mean errors | *SD* | Range |
|---|---|---|---|---|---|---|
| Motor | 21.80 | 7.49 | 10–33 | 91.76 | 26.30 | 40–146 |
| Verbal | 6.20 | 3.74 | 2–17 | 18.28 | 14.13 | 1–48 |

suggests a greater scope of application of theoretical explanations for serial order phenomena than heretofore suspected. Without concerning ourselves with this problem, we merely note that whatever principles are at work they are more readily displayed in the motor task because of the greater difficulty of this task and its high error score. It should be noted that one *S* learned the nonsense material with only one error which would provide no curve at all. The successful learning of the motor task by all *S*s suggests that it is a feasible nonverbal and presumably culture-free task that merits exploitation by those who need such an exercise. The use of mannequin displays as paired associated is, of course, a ready possibility. The stimulus side could consist of other mannequins, nonsense figures, or meaningful material, should that be of interest. An immediate prospect is comparisons of *S*s from different cultures in what appears to be a reasonable sample of associative learning where no *S*s would enjoy any special advantage.

## REFERENCES

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76** (1, Pt. 2).

SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale University, 1946.

(Received October 6, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 357–360

# ACOUSTIC SAVINGS FOR ITEMS FORGOTTEN FROM LONG-TERM MEMORY[1]

THOMAS O. NELSON[2]
*University of Washington*

ROBERT ROTHBART
*Stanford University*

A widely held view is that semantic, but not acoustic, information is stored in long-term memory (LTM). The present study explored the possibility that acoustic information might comprise part of the savings residual for items forgotten from LTM. Twenty undergraduate *Ss* originally learned (OL) a list of 24 number-word paired associates and 4 wk. later returned for a retention test and subsequent relearning (RL). During RL, the OL stimuli were paired with responses that were either homophones of OL responses or acoustically unrelated to the OL responses. For items incorrectly recalled on the retention test, RL performance was better for the homophone condition than for the acoustically unrelated condition, suggesting that acoustic information is stored in LTM and comprises part of the savings residual for items forgotten from LTM.

The savings effect (Ebbinghaus, 1885) refers to the established finding that the relearning of a list of verbal items is easier than the original learning of that same list. The present study investigated the savings effect within the framework of two-state memory (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965), where the two states are short-term memory (STM) and long-term memory (LTM). A paradigm has recently been devised (Nelson, 1971b) to investigate the savings effect for only those items which have been *forgotten* from LTM. In this paradigm, a list of items is originally learned (OL), followed by a long retention interval, followed by a retention test and subsequent relearning (RL). During RL, individual items either are changed or remain the same as during OL– savings is defined as the difference in RL performance between same and changed previously forgotten items. Prior research using this paradigm (Nelson, 1971a) has demonstrated considerable savings during RL, both for items that were not recalled on the retention test and for items that were not recognized on the retention test as having been previously learned. However, it is not sufficient to simply demonstrate that the savings residual can facilitate RL for forgotten items. One should also inquire into the nature of that savings residual, i.e., what properties does the savings residual have?

Structural models (e.g., Bower, 1967) have conceived of the memory trace as a cluster of features (e.g., semantic features, acoustic features), and the popular view (Adams, 1967; Baddeley, 1966; Dale & Baddeley, 1969; Norman, 1969) is that acoustic features are stored only in STM, while LTM stores only semantic features. However, some recent findings (Baddeley, 1970; Buschke & Lenon, 1969; Laurence, 1970; Wickens, Ory, & Graf, 1970) suggest that this feature–memory-state dichotomy may not be accurate: Retrieval from LTM may also involve acoustic information. If acoustic information is stored in LTM, this could have direct import for research on the nature of the savings residual. The present study explored the possibility that the savings residual for items forgotten from LTM is comprised of acoustic information.

## METHOD

*Design.*—Using a mixed-list design, the independent variable was the relationship between the OL response and the RL response. The list for a given *S* consisted of 24 paired associates, in which the stimuli were two-digit numbers of 1.36–1.99 association value (Battig & Spera, 1962) and the responses were common words with a frequency of 1–100 per million in the Thorndike-Lorge norms. The relationship between the OL response and the RL response

---

[1] This research was conducted while the first author was a National Science Foundation Postdoctoral Fellow at Stanford University, 1970–1971.

[2] Requests for reprints should be sent to Thomas O. Nelson, Department of Psychology, University of Washington, Seattle, Washington 98195.

was either (*a*) identity (e.g., 27-TACKS during OL, 27-TACKS during RL), (*b*) homophone (e.g., 27-TAX during OL, 27-TACKS during RL), or (*c*) unrelated control (e.g., 27-JURY during OL, 27-TACKS during RL). Because recent research (Cermak, Schnorr, & Buschke, 1970; Juola, Fischler, Wood, & Atkinson, 1971) has shown that LTM may retain visual (i.e., graphemic) information, the homophone condition was divided into visually similar homophones (e.g., PREY, PRAY) and visually dissimilar homophones (e.g., TAX, TACKS); the scheme used for this division took into account the number of common letters in the same serial positions, the number of common letters in different serial positions, and the length of the words. The resulting sets of visually similar and visually dissimilar homophones are shown in Table 1. For any given *S*, 6 of the 24 items were in the identity condition, 6 were visually similar homophones, 6 were visually dissimilar homophones, and the remaining 6 were unrelated control items; items were counterbalanced across *S*s so that each item appeared in the identity, homophone, and control conditions approximately equally often. In order to eliminate item bias during RL performance, the experimental manipulation occurred during OL rather than during RL. That is, all *S*s had the *same* RL list so that any differences in RL performance could be attributed to the residual effects of OL rather than to inherent item differences, except for the comparison between visually similar and visually dissimilar homophones, where, of necessity, different item pools had to be employed.

*Subjects.*—The *S*s were 20 Stanford undergraduates whose participation fulfilled a course requirement.

*Procedure.*—During OL, *S* first saw each of the 24 number-word pairs for 5 sec., then he shadowed numbers for 20 sec. to eliminate recall from STM (Glanzer & Cunitz, 1966; Postman & Phillips, 1965), and finally he had 10 sec. per item to spell aloud the correct response when he was presented with the stimulus number alone. To minimize overlearning, items correct on Trial *N* were eliminated from the study and test portions on Trial $N + 1$, until eventually every item had been correctly recalled once; this dropout procedure (Battig, 1965) was then repeated, beginning anew with the entire list, until all 24 items were correct on a single test trial. After attaining the OL criterion of one completely errorless test trial on all 24 items, *S* left the laboratory without being told that he would have a future retention test.

TABLE 1

ITEMS USED TO EXEMPLIFY VISUALLY SIMILAR AND VISUALLY DISSIMILAR HOMOPHONES

| Homophone pair | | | |
|---|---|---|---|
| Visually similar | | Visually dissimilar | |
| PREY | PRAY | TAX | TACKS |
| HEAL | HEEL | TOWED | TOAD |
| ALTER | ALTAR | VEIN | VANE |
| CREAK | CREEK | DOE | DOUGH |
| DUAL | DUEL | AUNT | ANT |
| PIER | PEER | INSIGHT | INCITE |
| SHEER | SHEAR | WINE | WHINE |
| LESSON | LESSEN | BANNED | BAND |
| STEAL | STEEL | AIL | ALE |
| POLE | POLL | IDOL | IDLE |
| FLEA | FLEE | WHALE | WAIL |
| PEEK | PEAK | PACT | PACKED |

Four weeks later *S* returned for the second session. First, there was a self-paced retention test where *S* had to spell aloud a response when each stimulus number was presented. Then *S* had one study-test trial on the RL list using the same timing and procedure as described above for OL.

## RESULTS AND DISCUSSION

The mean percentage of retention-test errors was 56.7%. This relatively small amount of forgetting after a 4-wk. retention interval is much lower than is usually found in studies of LTM but is typical (cf. Nelson, 1971a) of results obtained with a paradigm that (*a*) eliminates recall from STM during OL so that subsequent retention-test errors are due to forgetting from LTM rather than to a combination of forgetting from LTM and forgetting from STM, (*b*) uses unitary (e.g., word) rather than multicomponent (e.g., consonant trigram) responses, and (*c*) employs a self-paced retention test to ensure that *S* has enough time to recall difficult long-latency responses.

The data pertaining to the question about the nature of the savings residual come from the RL performance for items incorrectly recalled on the retention test, i.e., items forgotten from LTM. This RL performance, as a function of the items' OL–RL condition, is shown in Fig. 1. There is ample evidence for savings, since RL performance is clearly better for the identity condition than for the control condition. However, the most important finding is that there is considerable acoustic savings for items forgotten from LTM: A binomial test of the difference between proportions showed that RL performance is markedly better for the homophone conditions than for the control condition ($p <$ .001). Also, the two homophone conditions do not themselves differ significantly ($p >$ .50), and the nonsignificant difference is not even in the direction that would be expected if there were visual savings for items forgotten from LTM. Finally, the homophone conditions have poorer RL performance than the identity condition ($p < .001$), implying that the savings residual is composed of more than only acoustic information (e.g., perhaps some form of semantic information). Thus, the present findings confirm the hypothesis that the savings residual for items forgotten from LTM is partially comprised of acoustic information; however, no evidence was found for visual (graphemic) information in the savings residual.

One point about the present approach should be mentioned. It seems likely that the discovery of transfer gradients for items that are forgotten, as opposed to those that are remembered, will maximally expedite the formation of structural models of the memory trace (cf. Nelson, 1971b, pp. 574–575). This is because "what is learned" (i.e., the memory trace) is not the response word per se—which is known long before the college *S* begins the experiment—but rather, what is learned is recall of that word when provided with a particular stimulus cue (e.g., in learning JACK IS FUNNY the word FUNNY is not learned; rather, what is learned is the association between JACK and FUNNY). Therefore, in the case of transfer gradients for remembered items, it is always possible that *S* retrieves nonacoustic information (e.g., semantic information) about associations, applies the word name from his "mental dictionary," and only thereafter uses the word name as an acoustic code to produce the observed transfer effects. For example, given a stimulus cue such as JOHN LIKES THE ?, suppose that the correct OL response was the word meaning "a female deer"; after recalling DOE, one could make use of the acoustic information in the word name during a transfer task (e.g., to DOUGH). However, this clearly would not mean that the word name's acoustic information was ever stored as part of the association to the stimulus cue. Now consider a very different situation where the transfer effects are due to subthreshold information contained in forgotten items: If the stored association cannot be recalled, then acoustic information affecting transfer cannot be coming from the subsequent application of a word name; rather, that acoustic information must have comprised at least part of the memory trace which was formed during the original learning of the association. Only in this situation does it become reasonable to postulate acoustic information as one of the structural properties of the stored memory trace—unless the possibility of subsequent "coding" (after recall) can be eliminated, claims about the structure of the memory trace may be specious.

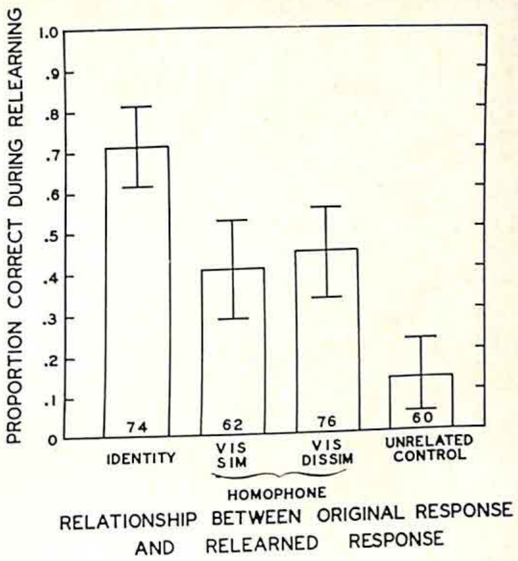


FIG. 1. Proportion of correct responses in relearning as a function of the relationship between the original response and the relearned response for items incorrectly recalled on the retention test. (Number inside bar is the *N* upon which the proportion is based; brackets show .05 level binomial confidence intervals.)

## REFERENCES

ADAMS, J. A. *Human memory.* New York: McGraw-Hill, 1967.

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

BADDELEY, A. D. The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology,* 1966, **18**, 302–309.

Baddeley, A. D. Effects of acoustic and semantic similarity on short-term paired-associate learning. *British Journal of Psychology*, 1970, **61**, 335–343.

Battig, W. F. Procedural problems in paired-associate learning research. *Psychonomic Monograph Supplements*, 1965, **1**, No. 1.

Battig, W. F., & Spera, A. Rated association values of numbers from 0–100. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 200–202.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Buschke, H., & Lenon, R. Encoding homophones and synonyms for verbal discrimination and recognition. *Psychonomic Science*, 1969, **14**, 269–270.

Cermak, G., Schnorr, J., & Buschke, H. Word recognition as a function of spelling direction during study and test. *Psychonomic Science*, 1970, **21**, 127–128.

Dale, H. C. A., & Baddeley, A. D. Acoustic similarity in long-term paired-associate learning. *Psychonomic Science*, 1969, **16**, 209–210.

Ebbinghaus, H. *Über das Gedächtnis*. Leipzig: Duncker, 1885.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971, **10**, 8–14.

Laurence, M. W. Role of homophones in transfer learning. *Journal of Experimental Psychology*, 1970, **86**, 1–7.

Nelson, T. O. Recognition and savings in long term memory: Related or independent? *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **5**, 15–16. (Summary) (a)

Nelson, T. O. Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 568–576. (b)

Norman, D. A. *Memory and attention*. New York: Wiley, 1969.

Postman, L., & Phillips, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 132–138.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Wickens, D. D., Ory, N. E., & Graf, S. A. Encoding by taxonomic and acoustic categories in long-term memory. *Journal of Experimental Psychology*, 1970, **84**, 462–469.

(Received October 7, 1971)

Journal of Experimental Psychology
1972, Vol. 93, No. 2, 361–366

# THE VON RESTORFF EFFECT IN SHORT-TERM MEMORY[1]

BARRY L. LIVELY[2]

*Bowdoin College*

An experiment is described in which the von Restorff effect was investigated in short-term memory using a probe technique. Isolation was produced at different list positions by embedding consonants among digits and digits among consonants. The predictions tested were those listed by S. E. Newman and E. Saltz in 1958; as predicted, it was found that performance on the item following the isolated item was superior to performance on an appropriate control item and the isolated item appeared less often as an intrusion. Contrary to predictions, the isolated item was not recalled with greater frequency than the control item, and performance on items other than the isolated item and its immediately following item was not superior to comparable portions of control lists. The results were interpreted in terms of stimulus recognition.

In the general case, the von Restorff effect refers to the enhanced recall of a perceptually or conceptually isolated item compared with the recall of a nonisolated control item. Wallace (1965), in a review of the literature, has shown the effect to be a pervasive phenomenon evident under several task conditions including serial and paired-associate learning tasks and free recall and immediate memory tasks.

Despite the long history of interest in the general nature of the effect, dating back to Calkins' (1896) investigations, little systematic investigation has been made of the von Restorff effect in short-term memory (STM). Cimbalo and his associates (Cimbalo, 1969; Cimbalo & Maloney, 1970; Cimbalo & Sielski, 1968), manipulating the position of the isolated item in a list, have reported enhanced recall of an item isolated by category (consonant vs. digit) or voice of the individual reading the list, but depressed recall of other items in the list in an immediate memory task involving recall of the whole list. Aside from these results, little other work of a systematic nature has been reported using STM techniques.

The experiment reported here was designed to provide data on the isolation effect as a function of serial position of the isolated item in an STM task. Isolation was produced at six locations in the list by embedding a consonant among digits and a digit among consonants. Memory for individual items was tested by the probe technique (Waugh & Norman, 1965) which tests recall of a single item by presenting the immediately preceding item. This technique was used to avoid the possible contaminating influence of requiring the recall of isolated and nonisolated items on the same trial.

The effects of isolation were examined in the framework of a generalization theory of the isolation effect. This theory was developed by Gibson (1940, 1942) and has evolved to the following set of predictions enunciated by Newman and Saltz (1958): (*a*) Performance on the isolated item as a response should be better than on a control item occupying the same serial position since it is less subject to response generalization from the other items in the list. (*b*) By the principle of stimulus generalization, the isolated item should be more differentiated as a stimulus, and performance on the item following the isolated item should be enhanced. (*c*) The presence of an isolated item will produce better performance on the remainder of the list since overall generalization tendencies are reduced within the list. (*d*) The isolated item should appear less often as an intrusion than the comparable control item.

---

[1] This research was supported by the Advanced Research Projects Agency, United States Department of Defense, and monitored by the United States Air Force Office of Scientific Research, under Contract AF 49(638)-1736 with the Human Performance Center, Department of Psychology, University of Michigan, and is based on the author's doctoral dissertation. The author is indebted to his advisor, Arthur W. Melton, for his thoughtful criticism and encouragement. Part of this report was given as a paper at a meeting of the Eastern Psychological Association, Philadelphia, April 1969.

[2] Requests for reprints should be sent to Barry L. Lively, Department of Psychology, Bowdoin College, Brunswick, Maine 04011.

Data relevant to these predictions were discussed by Wallace (1965). While all these predictions have not received general support in the literature, they provide a logical context for examination of possible effects.

## METHOD

*Subjects.*—The *S*s were 40 male volunteers drawn from a pool of paid *S*s at the University of Michigan.

*Procedure.*—The procedure on each trial consisted of presenting *S* with a string of 10 items at a rate of 1 item per $\frac{2}{3}$ sec. over a tape recorder, followed by *E* speaking 1 of the items (the probe) of the string. The *S* had 6 sec. to respond followed by a 4-sec. rest period; a warning signal preceded the next string by 2 sec. The successive stages of a trial were marked by clicks on the tape recorder. The task for *S* was to respond orally with the item following the probe item in the list and to give a confidence rating for his response. There were three available confidence ratings: LOW, which indicated a guess or very little confidence; MEDIUM, which meant greater confidence; and HIGH, which indicated that *S* was quite sure that his response was correct. A card containing the possible responses (consonants and digits) and confidence ratings was available to *S* at all times. The *S* was instructed that if the last item was presented by *E* as the probe, the first item of the string was the appropriate response.

Each *S* was tested in two sessions. In one session, *S* was given consonants isolated among digits (and all-digit control trials) and, in the other, digits isolated among consonants (and all-consonant control trials). The order of kinds of sessions was counterbalanced across *S*s. At the beginning of each session, *S* was given 14 practice trials on the material to be used followed by three blocks of 70 trials each. At the end of each block, a 4–5 min. break was given and *S* was informed of the number of correct responses he had made.

In order to maintain motivation throughout a session and to prevent *S* from hypothesizing about the relative importance of the various items, 1¢ was paid for each correct response in addition to the base pay of $1.25 per hour. The total time required for each *S* over the two sessions was approximately $3\frac{1}{2}$ hr.

Isolation was produced by inserting at Positions 1, 3, 5, 7, 9, or 10 either a single consonant among a set of nine digits or a single digit among nine consonants. Control trials with either homogeneous sets of digits or consonants were also given. Ten lists of each kind were presented in each block of 70 trials. The stimuli used were the digits from 0 to 9 and the consonants C, D, F, G, H, J, K, L, N, R. These were selected for relatively low rate of acoustic confusion (Conrad, 1964).

Within a trial block, no letter or digit was permitted to occupy the same serial position in successive strings, and the probe item was different from trial to trial. The frequency of usage of the nonisolated elements as probes was kept within the range of four–eight times within a block, while frequency of usage of isolated items as probes was maintained at six.

The ordering of kinds of lists was essentially random except that the number of intervening trials between successive exemplars of one kind of list could be no greater than 15 and no fewer than 1. In addition, all varieties were represented in the first and last 10 trials of the block.

For all *S*s, each serial position in each kind of list was tested once per block. For different *S*s, a different position was tested in a given list. The same lists were presented to all *S*s.

## RESULTS

The dependent variable analyzed was proportion or frequency of correct (and incorrect) responses at relevant serial positions under each condition of isolation. Although confidence data were collected, it is not reported here since no important trends were observed.

The data presented are organized according to the predictions made earlier. Inspection revealed no important differences between sessions; the analyses reported were performed on the combined data from both sessions. The data concerning the isolated item as a response are shown in Fig. 1 where isolated consonants (IC) are in Panel A and isolated digits (ID) are in Panel B. Nonisolated (control) digits are designated as NID and nonisolated (control) consonants as NIC.

No overall isolation effect is apparent in either Panel A or Panel B. A statistic derived for the combination of 2 × m contingency tables (Kincaid, 1962)[3] indicates no overall difference for either isolated consonants ($p > .75$) or for isolated digits

---

[3] This statistic represents a generalization of the method developed for dealing with 2 × 2 contingency tables (Cochran 1954). It is distributed approximately as chi-square. The analysis is similar to a fixed effects analysis of variance in that generality is limited by considering *S*s as individuals rather than random representatives of the general population.

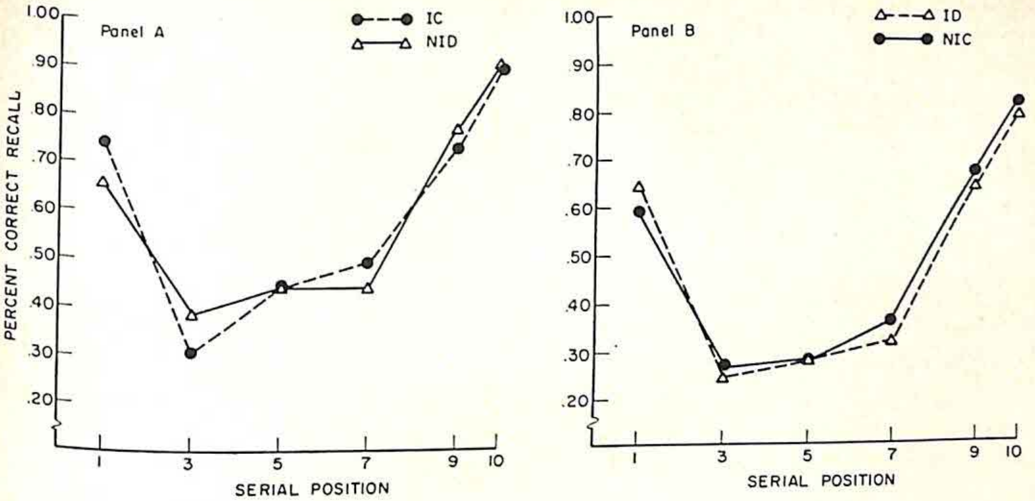


FIG. 1. Proportions of correct recall of an isolated item as a response to probe cue in the immediately preceding serial position, displayed for comparison with recall of a nonisolated item with the same context. (Panel A is for an isolated consonant in SP 1, 3, 5, 7, 9 or 10 of a string of digits and for nonisolated digits in those positions; Panel B is for an isolated digit in a string of consonants and for nonisolated consonant in the same position.)

($p > .50$). Performance was compared for the two conditions at each position, and in no case was a significant difference obtained ($p > .10$ in each comparison). Thus, the generalization theory's prediction that the isolated item would benefit in recall is not supported.

Comparisons between panels reveal that while control digits are recalled significantly better than control consonants ($p < .001$), isolated consonants are recalled with higher probability than isolated digits ($p < .001$). Comparisons of IC with NIC and ID with NID reveal a marked difference in performance with consonants benefiting from the digit context ($p < .001$) and digits suffering from the consonant context ($p < .001$), which suggests the determining role of the remaining items in the recall of the isolated item.

Figure 2 shows the effect of using the isolated item as stimulus. Panel A shows

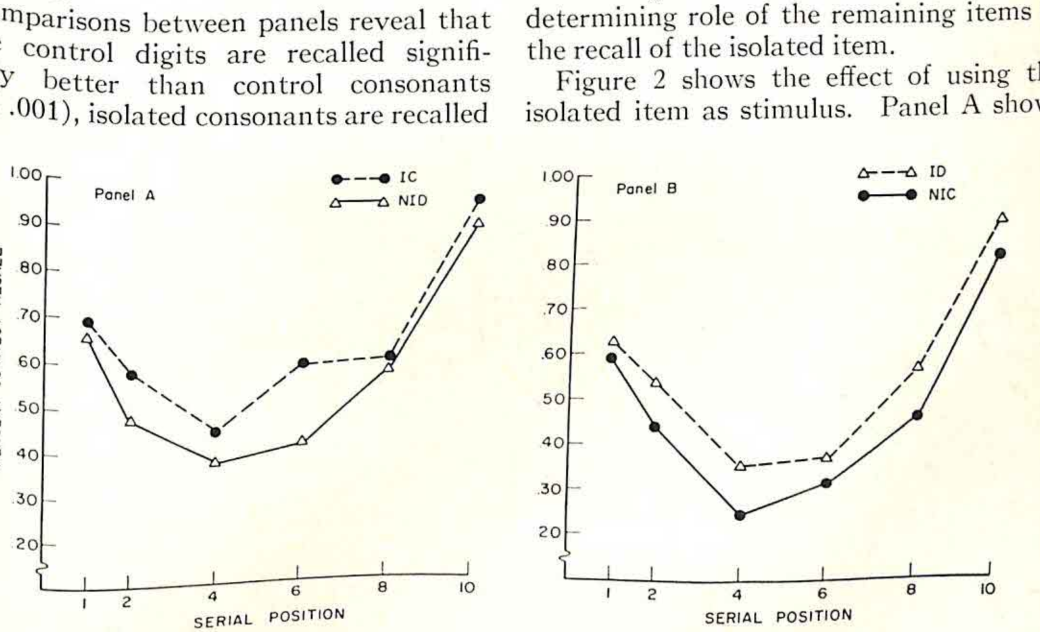


FIG. 2. Proportions of correct recall of the item immediately following the isolated item, when the isolated item is the probe cue, displayed for comparison with recall when a nonisolated item in the same context is used as the probe cue.

the recall of digits with either an isolated consonant or nonisolated digit serving as probe, while Panel B shows the recall of a consonant with either an isolated digit or nonisolated consonant serving as probe. Confirming the prediction, the overall effect was enhancement of the recall of the item following the isolated item, both for isolated consonants ($p < .005$) and isolated digits ($p < .001$).

These analyses include the data for the recall of the item in Serial Position 1, the probe for which was the last item of the list. Differential performance might not be expected on this item since probe and response were not contiguous in presentation. Inspection of Fig. 2 reveals only a slight effect at this position.

To test the generalization theory's prediction that the presence of an isolated item will produce better performance on the remainder of the list, overall recall scores for each list were tabulated, with scores for the isolated item and the following item removed. Contrary to this prediction, analysis by the Kincaid (1962) statistic revealed no significant difference for either digit or consonant lists ($p > .10$ in both cases).

The foregoing analysis was in terms of total correct responses in each condition exclusive of performance on the isolated item (or its control) and the item following it. The possibility that a facilitative or depressive effect is restricted to the immediate vicinity of the isolated item was checked in separate analyses of variance performed on (*a*) the item immediately preceding and (*b*) the item occurring two positions before the isolated and appropriate control item. The results were uniform in showing no effect of isolation on these positions. In all cases, the $F$ ratios were less than unity.

TABLE 1

TOTAL INTRUSIONS FROM POSITION (i) OF ISOLATED ITEM OR ITS CONTROL ITEM IN THE RECALL OF ALL OTHER TESTED ITEMS

| Position i | Cond. | | | |
|---|---|---|---|---|
| | IC | NID | ID | NIC |
| 1 | 6 | 29 | 21 | 34 |
| 3 | 36 | 34 | 38 | 32 |
| 5 | 45 | 52 | 46 | 62 |
| 7 | 67 | 83 | 66 | 89 |
| 9 | 74 | 76 | 82 | 105 |
| 10 | 24 | 19 | 30 | 45 |
| Total | 252 | 293 | 283 | 367 |

Note.—Abbreviations. IC = isolated consonants, NID = nonisolated digits, ID = isolated digits, NID = nonisolated digits.

Frequencies of intrusions for the various conditions appear in Table 1. Each entry represents the number of times an isolated item (or its control counterpart) from a given position appeared as an error when another item in the list was tested. The Kincaid (1962) statistic was computed for comparisons of the sets IC versus NID and ID versus NIC after adjustment was made for opportunity in each condition. The number of opportunities was defined as the number of errors at positions where the isolated item was neither probe nor correct response. As predicted, the isolated item appeared less often as an intrusion both for consonants ($p < .05$) and digits ($p < .001$).

## DISCUSSION

The present experiment was designed to test predictions about the von Restorff effect in STM which are derived from a simple generalization theory. As predicted, it was found that recall of the item following the isolated item was enhanced and the isolated item appeared less often as an intrusion than the control item. Contrary to predictions, recall of the isolated item was not superior to recall of a control item, and recall of the remainder of the list did not benefit by the presence of the isolated item. These results are at variance with those reported by Cimbalo and Sielski (1968) and Cimbalo (1969). Beyond noting the difference in technique used (probed recall of a single item vs. recall of the whole list) an attempt at reconciling the differences would not be profitable.

The work of Martin (1967a, 1967b) is suggestive in interpreting the lack of a response effect and the finding of a stimulus effect. In these studies, *S*s learned eight pairs of consonant trigram, single-digit combinations (e.g., XZJ-7) by a study-test method. During test trials, the stimuli used on study trials were intermixed among new stimuli. The *S*'s task during test trials was to indicate whether each stimulus was presented during study trials (OLD), or not (NEW), and then give a response of "1" through "8." It was found that correct responding was dependent on recognition of the stimulus as OLD and responses made following nonrecognition were at chance level.

In a later analysis of the same data, Martin (1968) found that among the NEW stimuli incorrectly identified as OLD (false positives), if a stimulus shared a first letter in common with an OLD stimulus, there was a greater than chance tendency to elicit the response associated with the OLD stimulus. Thus, where during study trials the first letter of each trigram serves as a discriminating feature, this same encoded feature may result in false recognition of a NEW stimulus as OLD. It follows that correct recognition is a function of the extent to which the stimuli of a list can be encoded in a different form, not only from one another but from possible foils. This hypothesis is closely related to the generalization theory if it is assumed that high recognition accuracy indicates low generalization, and low recognition accuracy indicates high generalization.

The recognition hypothesis may be applied to the interpretation of the present experiment. Since the isolated item differs from the remaining items of the list on a conceptual dimension, it would be expected that the encoded representation would reflect this difference. Consequently, if a consonant is embedded in a set of digits and the consonant is used as probe, it would be better recognized than a digit control item since the functional representation of the consonant has at least one feature not shared by the remaining items of the list, while the digit may not be so uniquely encoded. Since a probe item was always drawn from the presented list, identifying the probe as a consonant (whether it is recognized as the particular consonant presented or not) may be a sufficient cue for retrieval since that represents a sufficient encoding of the item with respect to its role as stimulus. No such simple encoding device is likely to be found which separates the control digit from the remaining items of the list. This does not mean that the perception of the isolated item was necessarily any more "vivid" or that it had any particular attention-getting value; it means only that there is a simpler way of separating an isolated item from other presented items than probably exists for the control item.

If it is assumed that, independent of the form of encoding, the associative strengths of the isolated item with the immediately preceding and following items are the same as for a comparable control item, the major findings of the present experiment would follow. The significant stimulus effect predicted by this interpretation has already been noted. The finding of a lack of difference in recall of the isolated and control item could be explained in the same way. The probes (stimuli) for these items are of the same conceptual class, and it would not be expected that one would be more efficiently encoded than the other, thus resulting in no difference in performance on the isolated and control items.

The unique encoding of the isolated item may also result in greater accuracy in recognition of the approximate position of that item in the list. Thus, if some other item is being probed for, the isolated item, as a possible response, could be rejected more readily than its control item, resulting in fewer intrusions of the isolated item. This was the case in the present experiment.

The recognition hypothesis can also account, in part, for the results of studies investigating the von Restorff effect in paired-associate learning. It has been found that isolation of a single stimulus term produced greater facilitation than isolation of a single response (Erickson, 1965; Kimble & Dufort, 1955), while isolation of two stimuli in the same list produced slightly, but not significantly, greater facilitation than isolation of two responses (Nachmias, Gleitman, & McKenna, 1961). This seems to imply that, as the number of isolated stimuli increases, the benefit accruing to enhanced recognition of these stimuli decreases, which is in line with the recognition hypothesis. The recognition hypothesis does not, however, account for the significant effect of response isolation. This effect may be due to enhanced association formation (Horowitz, 1962) or increased response availability (Saltz & Newman, 1959).

## REFERENCES

CALKINS, M. W. Association: An essay analytic and experimental. *Psychological Review*, 1896 (Monograph Supplement No. 2).

CIMBALO, R. S. Short-term memory: The effects of task-relevant isolation. Paper read at the meeting of the Eastern Psychological Association, Philadelphia, April 1969.

CIMBALO, R. S., & MALONEY, M. S. Short-term memory: Effects of type of trigram isolated and position of isolation. *Psychonomic Science*, 1970, 18, 77–78.

Cimbalo, R. S., & Sielski, R. L. Short-term memory: The effects of item isolation upon performance. Paper read at the meeting of the Eastern Psychological Association, Washington, D. C., 1968.

Cochran, W. G. Some methods for strengthening the common $\chi^2$ tests. *Biometrics*, 1954, **10**, 417–451.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

Erickson, R. L. Differential effects of stimulus and response isolation in paired-associate learning. *Journal of Experimental Psychology*, 1965, **69**, 317–323.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, **47**, 196–229.

Gibson, E. J. Intra-list generalization as a factor in verbal learning. *Journal of Experimental Psychology*, 1942, **30**, 185–200.

Horowitz, L. M. Associative matching and intralist similarity. *Psychological Reports*, 1962, **10**, 751–757.

Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *Journal of Experimental Psychology*, 1955, **50**, 361–368.

Kincaid, W. M. The combination of 2 X m contingency tables. *Biomertics*, 1962, **18**, 224–228.

Martin, E. Relation between stimulus recognition and paired-associate learning. *Journal of Experimental Psychology*, 1967, **74**, 500–505. (a)

Martin, E. Stimulus recognition in aural paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 272–276. (b)

Martin, E. Supplementary report: Recognition and correct responding medicated by first letter of trigram stimuli. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 703–706.

Nachmias, J., Gleitman, H., & McKenna, V. The effect of isolation of stimuli and responses in paired-associates. *American Journal of Psychology*, 1961, **74**, 452–456.

Newman, S. E., & Saltz, E. Isolation effects: Stimulus and response generalization as explanatory concepts. *Journal of Experimental Psychology*, 1958, **55**, 467–472.

Saltz, E., & Newman, S. E. The von Restorff isolation effect: Test of the intralist association assumption. *Journal of Experimental Psychology*, 1959, **58**, 445–451.

Wallace, W. P. Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, 1965, **63**, 410–424.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

(Received October 12, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 367–372

# THE LATENCY OPERATING CHARACTERISTIC: II.

# EFFECTS OF VISUAL STIMULUS INTENSITY ON CHOICE REACTION TIME[1]

JOSEPH S. LAPPIN[2] AND KENNETH DISCH

*Vanderbilt University*

The purpose of this experiment was to study the effect of stimulus intensity on the latency operating characteristic (LOC), which is the trade-off relation between speed and accuracy. The stimuli were presented at three intensities that differed by equal intervals on a logarithmic scale. Both the slope and intercept of a linear relation between reaction time (RT) and discrimination accuracy (*d′*) were found to vary with stimulus intensity: As intensity increased, the slope of the LOC increased and the intercept decreased. The greatest differences were between the lower two intensities. The findings contribute discriminant validity to the LOC as a measure of a perceptual process.

The "latency operating characteristic" (LOC) refers to the functional relation between RT and discrimination accuracy. The usefulness of this measure lies in its potential as a measure of the perceptual process that remains invariant under changes in decision strategies that determine when the response to a given percept will be emitted. Our initial report (Lappin & Disch, 1972) provided evidence for the convergent validity of the LOC by showing its invariance under changes in stimulus and response probability. The LOC was described by the slope and intercept of a straight line relating reaction time (RT) and $d'$ (or one of several alternative measures of discrimination accuracy).

The purpose of the present experiment is to study the discriminant validity of the LOC—by determining the effects of stimulus intensity upon the slope and intercept of the LOC. If the LOC is in fact a measure of the perceptual process, then the parameters of the LOC should vary as a function of stimulus intensity.

An inverse relation between stimulus intensity and RT has been repeatedly demonstrated, but the details of the relationship are uncertain: Decisional processes appear to contribute a major influence on the relation between intensity and RT. Grice (1968), Kohfeld (1969), and Murray (1970) have demonstrated that the relationship depends upon such nonperceptual variables as response incentives, amount of practice, and the distribution of intensity values over trials. The LOC offers a possible means for isolating the logically distinct perceptual and decisional processes underlying performance in these RT tasks.

Another potential advantage of the LOC methodology for assessing the effects of stimulus intensity pertains to the use of a different set of performance measures—the slope and intercept of the LOC—in place of the conventional statistics for describing distributions of correct RT responses. Little is known, however, about the dependence of the slope and intercept of the LOC upon such stimulus variables as intensity. Grice (1968) and McGill (1967) have suggested some reasons for hypothesizing that the rate of gain of sensory information should increase with stimulus intensity; and this effect might be measurable by the slope of the LOC, which reflects the rate of gain in discrimination accuracy with increasing RT. Pachella and Fisher (1969), however, have obtained data indicating that stimulus intensity affects only the intercept but not the slope of the LOC. The intercept of the LOC, which reflects the lag in RT before discrimination begins to rise above a chance level, would seem to correspond to the effect that is implied in the conventional use of the mean RT statistic as a measure of the speed of the perceptual process. The intercept of the LOC might thus also be expected to decrease as the stimulus intensity is increased.

---

[1] This study was supported in part by funds from the Graduate Research Council at Vanderbilt. Some of the results were presented at the Psychonomics Society meetings, November 1970.

[2] Requests for reprints should be sent to Joseph S. Lappin, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37204.

## METHOD

*Stimuli.*—Most of the methodological details of the present experiment were identical to those described in the initial report. Two easily discriminable stimuli (to the right or left of a fixation point) were mapped onto two highly compatible responses (right hand vs. left hand). The stimuli were luminous circles, 1° in diameter and 40′ off center to the right or left of the fixation point, which was centered in a $5\frac{1}{2}°$ square luminous background. Three stimulus intensities were employed: 6.0, .6, and .06 ftl. (controlled by neutral density filters). Luminances were measured with a Spectra Spot photometer. These stimulus luminances were added to a continually present background of 2.5 ftl., and remained in view until *S* responded. The lowest intensity was just barely visible—usually only fragments of the complete circle were visible but the stimulus was never missed entirely. The stimuli were presented and controlled by the same tachistoscopic apparatus as described in the previous report.

*Procedure.*—A speed-accuracy trade off was obtained by asking well-practiced *S*s to respond at a speed that would result in an error rate of about 25%. The *S*s were instructed to avoid guessing but to adjust their RTs from trial to trial in order to maintain an average speed corresponding to the desired error rate.

All three intensities were presented in each session—in blocks of 130 trials, consisting of 30 practice trials and 100 experimental trials. Each *S* served for six experimental sessions, preceded by three practice sessions. The order of the three stimulus intensities was counterbalanced across *S*s and sessions.

Other aspects of the procedure duplicated those of the previous (Lappin & Disch, 1972) experiment.

*Subjects.*—Five Vanderbilt students served as paid volunteers. One *S* was female. None had served in the previous experiment.

## RESULTS

Measures of speed and accuracy were calculated from the data by the same analytic procedure described in the previous article (Lappin & Disch, 1972). The RTs were first rank ordered and then partitioned into a series of sets of 75 trials. The median RT was used as the measure of speed for each set of responses: $d'$ (assuming normal discriminal distributions with equal variance) was the measure of discrimination accuracy that was calculated from the $2 \times 2$ confusion matrix constructed from each successive set of 75 responses. Eight pairs of speed and accuracy measures were thus obtained from the data pooled across sessions for each *S* and condition. The number of trials in each point on the LOC was somewhat arbitrarily set at 75, as a compromise between maximizing the reliability of the $d'$ measures and maximizing the number of points from which the LOC parameters were computed. Some of the variability in the obtained points reflects the unreliability of the $d'$ measures with sets of 75 responses.

The LOC was described by a pair of straight lines that were fit to this set of points by the procedure described by Bogartz (1968). One line described performance at a chance level of accuracy, $d' = 0$, for $\text{RT} < k$; the other line described the region of the speed-accuracy trade off, $d' = m(\text{RT} - k)$, for $\text{RT} \geq k$. The slope, $m$, and the intercept, $k$, of the second line provided the parameters to describe the LOC. The ranges of RT values to which each of the line segments was fit were selected so as to minimize the sum of squared deviations from the lines, under the restrictions that (*a*) the intercept $k$ did not lie within the range of RT values to which the line segment $d' = 0$ was being fit, and (*b*) all RTs were within the range where perfect discrimination did not occur. (One exception to this rule was permitted for the data of *S* S. L. in the .6-ftl. intensity condition, where just 2% more variance would have been described by an LOC with slope and intercept parameters quite different from those obtained for other sets of data—$m = 5.20$ instead of 4.06, and $k = 171$ instead of 158. The former parameters were obtained by placing one more point on the line at $d' = 0$ rather than on the trade-off segment of the LOC.)

The results of this analysis are given in Table 1 and Fig. 1. Table 1 gives the slopes, $m$, and intercepts, $k$, for the equation $d' = m\,(\text{RT} - k)/100$ describing the LOC for each *S* and condition, and also

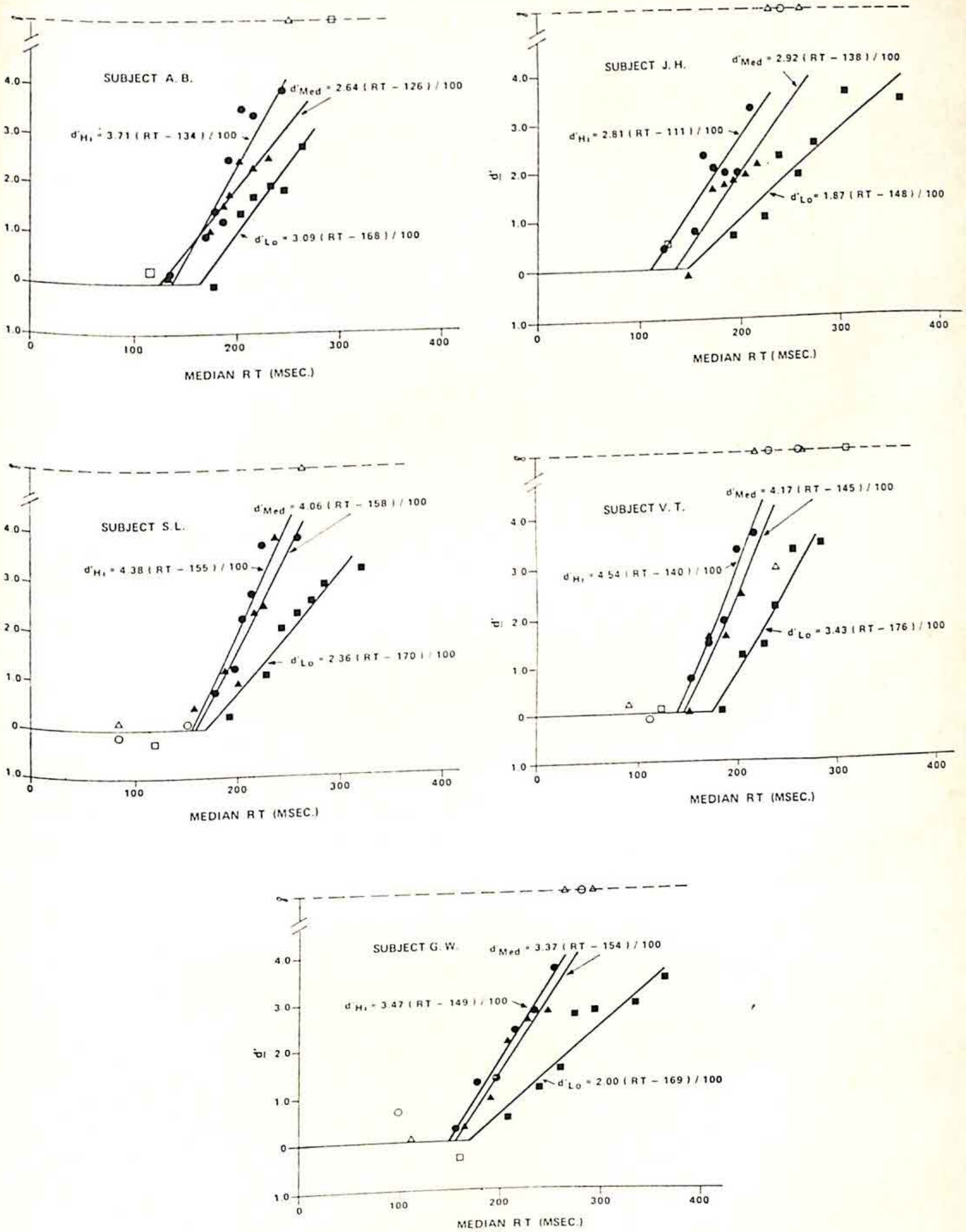


FIG. 1. LOC functions for each *S* and condition. (The circles are from the 6.0 ftl. condition, triangles from the .6 ftl. condition, and squares from the .06 ftl. condition. The lines given by the equations were fit to filled data points.)

the percentage of the variance in $d'$ values that is accounted for by each of these LOCs. As is apparent, the functional relation between speed and accuracy is adequately described by an LOC consisting of two linear segments expressing $d'$ as a function of RT. It is also apparent that increases in stimulus intensity tend to

TABLE 1

THE SLOPES, $m$, AND INTERCEPTS, $k$, FOR THE SEGMENT OF THE LOC DESCRIBED BY THE EQUATION $d' = m(\mathrm{RT} - k)/100$, FOR EACH $S$ AND CONDITION; AND THE VARIANCE IN $d'$ ACCOUNTED FOR BY BOTH SEGMENTS OF THE LOC

| $S_s$ | Slope | | | Intercept | | | Percentage of variance | | |
|---|---|---|---|---|---|---|---|---|---|
| | 6.0 | .6 | .06 | 6.0 | .6 | .06 | 6.0 | .6 | .06 |
| A. B. | 3.71 | 2.64 | 3.09 | 134 | 128 | 168 | .825 | .930 | .907 |
| J. H. | 2.81 | 2.92 | 1.87 | 111 | 138 | 148 | .787 | .808 | .872 |
| S. L. | 4.40 | 4.06 | 2.36 | 155 | 158 | 170 | .959 | .872 | .853 |
| V. T. | 4.54 | 4.17 | 3.43 | 140 | 145 | 176 | .977 | .918 | .959 |
| G. W. | 3.47 | 3.37 | 2.00 | 149 | 154 | 169 | .937 | .958 | .948 |
| $\bar{X}$ | 3.79 | 3.43 | 2.55 | 138 | 144 | 166 | .897 | .897 | .928 |

produce increases in the slope and decreases in the intercept of the LOC. The largest effects of stimulus intensity are manifest in comparisons between the .6- and .06-ftl. conditions.

Statistical analyses of the effects of stimulus intensity were performed separately for the slope and intercept parameters. Significant effects were obtained in both cases: for the slopes, $F(2, 8) = 10.3$, $p < .01$; for the intercepts, $F(2, 8) = 14.5$, $p < .01$. Newman-Keuls post hoc comparisons among the means for the three intensity conditions revealed no significant differences in the slopes for either the 6.0 versus .6 or .6 versus .06 conditions, but a significant difference in the intercepts of the .6 and .06 conditions, $p < .01$. Although performance was reliabley better in the 6.0 than the .6 condition (see data below), there is some uncertainty about whether the difference is more accurately reflected in the slopes, the intercepts, or in both of these parameters.

To obtain a different view of these data we also computed the discrimination accuracy for each of the conditions for RTs within given bands of time. Specifically, we calculated $d'$ values for each of five successive 20-msec. bands of RTs from 141 to 240 msec., for each $S$ and condition. The average values for each condition are given in Table 2. Statistical analyses for these data used Friedman's two-way analysis of variance for ordinal data (see Siegel, 1956) applied separately to the data for each time band. The resulting $X_r^2$ ($k = 3$, $n = 5$) statistics were 1.6, 5.2, 10.0 (the maximum value obtainable, $p < .01$), 10.0 ($p < .01$), and 7.6 ($p < .05$), respectively, for each of the five successive time bands. As is apparent from these statistics and also from Table 2, the separation between conditions increases with time, corroborating the increase in slope of the LOC as a function of intensity. Applying the same analyses to the data for each $S$, treating the time bands as independent replications, yielded $X_r^2$ ($k = 3$, $n = 5$) values of 6.4 ($p < .05$), 10.0 ($p < .01$), 3.6, 3.6, and 6.4 ($p < .05$) for $S$s A. B., J. H., S. L., V. T., and G. W., respectively. Comparing the discrimination accuracy for each pair of intensity conditions at all 25 such data points (5 $S$s × 5 Time Bands) also reveals consistent differences in performance among the three intensity conditions: Discrimination was more accurate for the 6.0 than for the .6 condition in 18 of the 25 cases ($p < .05$, by the binomial distribution), more accurate for the .6 than the .06 condition in 23 cases ($p < .01$), and more accurate in the 6.0 than the .06 condi-

TABLE 2

AVERAGE VALUES OF $d'$ FOR FIVE SUCCESSIVE BANDS OF REACTION TIME

| Range of RTs | Stimulus intensity (ftl.) | | |
|---|---|---|---|
| | 6.0 | .6 | .06 |
| 141–160 | .60 | .40 | .29 |
| 161–180 | 1.14 | 1.02 | .13 |
| 181–200 | 1.80 | 1.62 | .47 |
| 201–220 | 2.76 | 2.00 | .90 |
| 221–240 | 3.53 | 3.04 | 1.55 |

tion in 23 cases ($p < .01$). Thus, when speed is held constant, the accuracy of discrimination increases with stimulus intensity.

One of the issues examined in our initial report (Lappin & Disch, 1972) concerned the selection of a measure of accuracy that would be linearly related to RT. Although a good fit was found for the $d'$ measure, the data could not clearly justify the selection of this measure over even the algebraically distinct measure $(d')^2$, that is derived from a sequential sampling model for RT. We considered this issue again with the present data, and again with unsatisfying results. For the 15 independent sets of data (five *S*s by three conditions), the $(d')^2$ measure gave a better linear fit in eight of the comparisons with the $d'$ measure, and the average amount of variance accounted for was quite similar for the two measures. This result clearly indicates that the $d'$ measure which we have chosen to use here is not unique in its description of these data. A variety of other models and measures could also be employed (see Lappin & Disch, 1972).

## DISCUSSION

The results of this experiment support the hypothesis that the slope and intercept of the LOC reflect the temporal characteristics of the acquisition of perceptual information. On the premise that stimulus intensity affects the speed of some aspect of a perceptual process (as opposed to decision and response processes), then the finding that the slope and intercept of the LOC are dependent upon stimulus intensity indicates that these parameters measure a perceptual process. This result contributes evidence for the discriminant validity of the LOC, combining with the evidence for convergent validity provided in our first report (Lappin & Disch, 1972). Presumably, the slope of the LOC (the rate of increase in discrimination accuracy as a function of RT) reflects the rate of gain of perceptual information, and the intercept of the LOC (the RT at which discrimination accuracy begins to rise above chance) apparently reflects a minimum lag between the presentation of a stimulus to *S*'s retina and the detection of the resultant neural signals in some central data-processing and decision-making system.

The relation between stimulus intensity and the slope of the LOC is consistent in several ways with Grice's (1968) suggestions about the temporal characteristics of the perceptual processes underlying performance in RT tasks. In the first place, the finding that the slope increases with stimulus intensity is compatible with Grice's model for the effects of stimulus intensity on simple RT. According to Grice's model, an increase in stimulus intensity produces an increase in the rate of the elicited neural activity, so that the differences in neural activity produced by various intensities increase with time. The present experiment provides a more direct confirmation of this model by examining the relation between speed and accuracy in a choice RT task. The LOCs recovered from these data correspond rather well with the hypothetical functions presented in Grice's paper.

An additional correspondence with Grice's (1968) model concerns the mathematical description of the LOC function relating speed and accuracy. Following McGill's (1967) formulation, Grice suggested that decisions to respond in certain RT tasks are based upon counts of neural impulses and that the number of elicited neural impulses increases as a linear function of time. One implication of this theory for the present experiment is that $(d')^2$ should be approximately linearly related to RT: suppose that (*a*) the stimulus acts to increase the firing rate from a value $\gamma$ to a new level $\alpha\gamma$ in neurons associated with that particular stimulus presentation, (*b*) the right or left response is selected by comparing counts of the total number of impulses associated with the two alternative stimulus positions and choosing the response associated with the larger count, and (*c*) the count at a point $t$ msec. from the initial registration of a stimulus is an approximately normally distributed random variable with expected value $\gamma t$ for no stimulus presentation and $\alpha\gamma t$ given a stimulus presentation and with standard deviation $\sigma t^{\frac{1}{2}}$. Following the signal detection theoretic representation, we may relate the frequencies in the confusion matrix to areas above and below criterion in two probability distributions of impulse counts for the two stimulus alternatives by the standardized difference in their average impulse counts:

$$\begin{aligned} d_{t'} &= (\alpha\gamma t - \gamma t)/\sigma t^{\frac{1}{2}} \\ &= t^{\frac{1}{2}}(\alpha - 1)(\gamma/\sigma) \qquad [1] \\ (d_{t'})^2 &= t[(\alpha - 1)(\gamma/\sigma)]^2 \end{aligned}$$

where $t = \mathrm{RT} - k$, in the equation $d' =$

$m(\text{RT} - k)/100$. McGill (1967) has shown how such a model can account for a variety of results in auditory detection, and the present data are not clearly inconsistent with that model. However, it must be emphasized that these data can also be described by a number of other distinctly different models—e.g., the additive model that predicts a linear relation between RT and $d'$ (see Lappin & Disch, 1972).

The dependence of the slope of the LOC upon stimulus intensity appears to be inconsistent with the findings of Pachella and Fisher (1969), who obtained an effect of stimulus intensity upon the intercept but not the slope of the LOC. This inconsistency is, though, probably due to a difference in the discrimination tasks employed in their experiment and ours. Whereas our *S*s were asked to discriminate between two distinctly different positions, their *S*s were required to discriminate among 10 alternative stimulus positions under conditions that resulted in errors for RTs of nearly 1.0 sec. Their task would seem likely to have reflected the operation of more cognitive processes, serving to compare the stimuli with a memory of the alternative positions. Presumably, our task depended on the operation of a more peripheral perceptual processing in which stimulus intensity plays a more prominent role.

The inverse relation between intensity and the intercept of the LOC seems to reflect the operation of an initial temporal integration process that accumulates stimulus information until some threshold has been surpassed. Given the very dim and incomplete appearance of the stimuli in the low-intensity condition, the detection of these stimuli probably required such temporal integration. The greatest differences in intercepts of the LOCs occurred between the medium-(.6 ftl.) and low-(.06 ftl.) intensity conditions, but it is questionable whether there is any reliable difference between the high-(6.0 ftl.) and medium-intensity conditions. Although the intercepts were lower for the high- than for the medium-intensity condition in four out of five cases, the differences were very small. Interestingly, the subsequent rate of gain of perceptual information—as reflected in the slope of the LOC—seems to proceed at a low rate following such integration.

A moot question concerns whether the slopes and intercepts of the LOCs provide measures of two independent processes or whether they reflect correlated aspects of the same underlying process. The effects of stimulus intensity produced correlated changes in both parameters in this experiment, suggesting the possibility that these parameters do not represent independent component processes. One wonders, though, whether other stimulus variables would also produce correlated changes in these parameters. A full and adequate interpretation of the slope and intercept parameters that describe the LOC will require more extensive evidence on the dependence of these parameters on other stimulus and task variables.

## REFERENCES

Bogartz, R. S. A least squares method for fitting intercepting line segments to a set of data points. *Psychological Bulletin*, 1968, **70**, 749–755.

Grice, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, **75**, 359–373.

Kohfeld, D. L. Effects of the intensity of auditory and visual ready signals on simple reaction time. *Journal of Experimental Psychology*, 1969, **82**, 88–95.

Lappin, J. S., & Disch, K. The latency operating characteristic: I. The effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology*, 1972, **92**, 419–427.

McGill, W. J. Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology*, 1967, **4**, 351–376.

Murray, H. G. Stimulus intensity and reaction time: Evaluation of a decision-theory model. *Journal of Experimental Psychology*, 1970, **84**, 383–391.

Pachella, R. G., & Fisher, D. F. Effect of stimulus degradation and similarity on the trade-off between speed and accuracy in absolute judgments. *Journal of Experimental Psychology*, 1969, **81**, 7–9.

Siegel, S. *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received October 12, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 373–378

# SEQUENTIAL EFFECTS OF STIMULUS PROBABILITY AND PREDICTION OUTCOME ON CHOICE REACTION TIME[1]

CHARLES P. WHITMAN AND E. SCOTT GELLER[2]

*Virginia Polytechnic Institute and State University*

Second-order effects of stimulus probability and prediction outcome (PO) on choice reaction time (RT) were studied in a two-stimulus, two-response paradigm. Choice RT was shorter following the more probable stimulus and correctly predicted (CP) events and was longest following an incorrectly predicted (IP) repetition of the less probable stimulus. Choice RT to the more probable alternative was not affected by the preceding stimulus; RT to the less probable stimulus was faster to repetitions when the stimulus was CP but faster to alternations when it was IP. Preceding probability and preceding PO interacted with current PO; i.e., RT to CP events was influenced by preceding PO but not by preceding probability; RT to IP events was differentially affected by both preceding PO and preceding probability. A fourth-order analysis of PO effects indicated a reliable third-order effect of PO on choice RT. The implications of the findings for expectancy theory are discussed.

As reviewed by Smith (1968), choice reaction time (RT) experiments which examined stimulus probability and prediction outcome (PO) effects have consistently found that *S*s react faster to the more frequently occurring stimulus alternative and to correctly predicted (CP) stimuli. Recent choice RT studies which investigated the effects of probability and PO within the same experiment found that effects of stimulus probability were independent of the effects of prediction outcome (Geller & Pitz, 1970; Geller & Whitman, 1971; Hinrichs & Craft, 1971). That is, when the choice RTs were dichotomized into CP and IP categories, *S*s reacted faster to the more probable stimulus alternative. The recent research of Geller, Whitman, Wrenn, & Shipley (1971) not only replicated the finding that stimulus probability and PO are independent determinants of choice RT, but also demonstrated that regardless of the PO, *S*s respond faster to stimulus repetitions than stimulus alternations in a discrete choice reaction task.

Whitman and Geller (1971, 1972) found that the PO on trials prior to a given stimulus presentation reliably influenced *S*s' speed to identify that stimulus. The consistent finding of these studies was that choice RTs to CP stimuli were faster if the preceding PO had been correct rather than incorrect. In a relatively incompatible task, a correct PO preceding a choice RT to an IP event also facilitated the reaction; however, in a highly compatible situation a preceding correct PO inhibited choice RTs to IP events (Whitman & Geller, 1972). A shortcoming of these studies was that the two stimulus alternatives were not equiprobable, and an insufficient number of trials prevented an examination of the influence of stimulus probability upon preceding PO effects. An analysis of preceding PO effects which does not first categorize trials according to stimulus probability may confound the effects of stimulus probability and preceding PO if the frequency of a particular stimulus varies between the categories of the analysis. In fact, depending on the degree to which the frequency differences influenced *S*s' predictions, from three to five times as many of the CP events would be the more

---

[1] Supported by The Small Grants Program of the College of Arts and Sciences, Virginia Polytechnic Institute and State University. This research was facilitated by college gift equipment contributed by Western Electric Company and the Chesapeake and Potomac Telephone Company. Special thanks are due to Richard F. Wrenn for constructing the logic equipment.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

probable stimulus when the stimuli vary on a 70/30 schedule (cf. Geller et al., 1971). Hence, so that the effects of current and preceding PO could be studied independently of current and preceding stimulus probability, *S*s in the present study received a large number of stimulus presentations.

A parameter of the preceding PO effect which has yet to be explored is the order of the sequential effect of PO (i.e., the number of POs preceding a given stimulus which significantly influences the choice RT to that stimulus). Whitman and Geller (1971, 1972) studied preceding PO effects by analyzing the influence on choice RT of runs of correct or incorrect POs preceding a correct or an incorrect prediction and determined second-order effects of PO. However, since those combinations of POs preceding a given trial which included both correct and incorrect predictions were excluded from the analysis, third-order effects of PO on choice RT were not determined. More specifically, allowing C and I to represent correct and incorrect POs in a sequence, these studies included categories IIC, CCC, CCI, and III, but not categories CIC, ICC, CII, nor ICI. Consequently, the present study considered the effects of all possible combinations of POs preceding choice RTs by using a tree analysis analogous to that used by Remington (1969) to study stimulus pattern effects on choice RT. The most remote of the preceding POs which has a reliable influence on the current reaction latency will determine the order of the sequential effect of PO. Even though the number of event categories increases exponentially as the number of events in the sequential analysis is increased, the large number of choice RTs collected in the present study made it possible to evaluate fourth-order effects of PO.

## METHOD

*Subjects.*—Twenty *S*s (14 males, 6 females) participated in four individual 1-hr. sessions on consecutive days. All *S*s were enrolled in psychology classes at Virginia Polytechnic Institute and State University and had not previously participated in a choice RT study.

*Apparatus and procedure.*—The two stimuli alternatives were the symbols $\underline{\cup}$ and $\underline{\cap}$ (verbally labeled "up" and "down," respectively). The $\underline{\cup}$ occurred on 70% of the trials and $\underline{\cap}$ occurred on the remaining 30%. These stimuli were presented by a 2 × 5 cm. digital readout (Allied No. MSM5A-60D7711). The identification response was made by pulling one of two microswitches mounted in spatially separated handles. Ten *S*s pulled the right switch to identify a $\underline{\cup}$ and the left to identify $\underline{\cap}$. The remaining 10 *S*s made a left trigger response to a $\underline{\cup}$ and a right response to a $\underline{\cap}$.

Each *S* received a total of 1,200 trials, 300 trials during each of the four sessions. The stimuli were randomized over blocks of 300 with the limitations that runs of the less probable stimulus did not exceed three and runs of the more probable stimulus did not exceed seven. The instructions urged *S*s to give thought to their predictions and emphasized the importance of both speed and accuracy in their stimulus identifications. The *S*s were not informed of the relative frequency of either stimulus but were told that the sequence was predetermined by a punched paper tape.

The sequence of events on each trial included the following: *S*s' verbal prediction, a brief warning buzzer, a random time interval ranging between .5 and 2 sec., a stimulus presentation, and *S*s' identification response, which turned off the stimulus and stopped a Hunter Klockcounter. The task was *S* paced with an interstimulus interval of 6 to 10 sec. Errors (i.e., anticipatory responses or incorrect identifications) occurred on less than 5% of the trials for all *S*s and were not considered in the analysis.

## RESULTS

*Prediction outcome and stimulus probability effects.*—To study the effects of stimulus probability and PO on choice RT, a 16-category analysis of the latencies on Trials 2 through 1,200 was employed. These categories were determined by the following four variables of two levels each: *current probability* (more or less frequent), *current PO* (CP or IP), *preceding probability* (more or less frequent), and *preceding PO* (CP or IP). The choice RT mean for each category indicated that reactions to CP events were markedly shorter than reactions to IP events (i.e., comparison between dotted and solid lines in Fig. 1). Moreover, latencies to the more probable stimulus, M, were consistently shorter than reactions to the less probable stimulus, L, (i.e., comparison between Panel 1 and Panel 2 in Fig. 1).

The overall ANOVA, a 2 (Current Probability) × 2 (Preceding Probability) × 2 (Current PO) × 2 (Preceding PO) design,

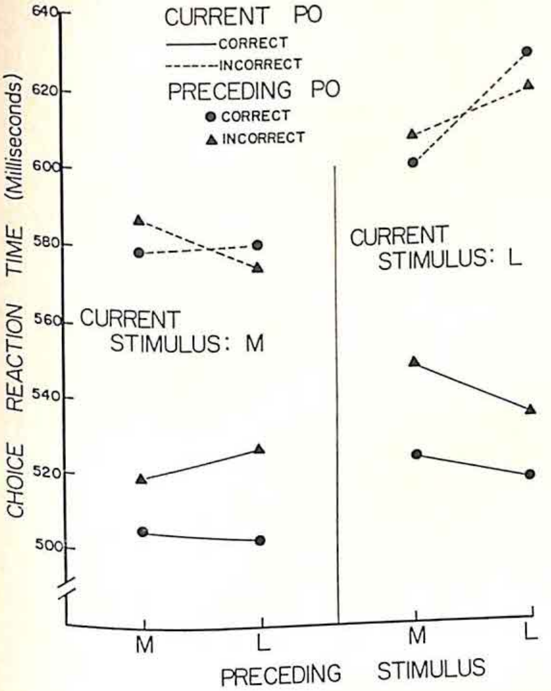


Fig. 1. Choice reaction time as a function of stimulus probability and prediction outcome.

indicated three significant main effects and five significant interactions. The main effects included that of current probability, $F$ (1, 19) = 36.56, indicating that *S*s reacted reliably faster to the more probable stimulus; that of current PO, $F$ (1, 19) = 36.95, demonstrating that *S*s reacted significantly faster to CP stimuli; and that of preceding PO, $F$ (1, 19) = 16.82, all $p$s $< .001$, showing that choice RT was significantly influenced by preceding PO. The effect of preceding probability was not reliable, $F$ (1, 19) = .29. The Current Probability × Current PO interaction, $F$ (1, 19) = 5.53, $p < .05$, was largely due to the greater effect of current PO when the current stimulus was L. In addition, the Preceding Probability × Current PO interaction, $F$ (1, 19) = 4.61, $p < .05$, indicated a greater effect of current PO when the preceding stimulus was L. However, the influence of preceding probability on current PO effects was minimal when the current stimulus was M (while it was quite robust when the current stimulus was L). Consequently, the third-order interaction, Current Probability × Preceding Probability × Current PO, was significant, $F$ (1, 19) = 11.27, $p < .005$.

A significant Current PO × Preceding PO interaction, $F$ (1, 19) = 16.82, $p < .001$, is accounted for by the greater effect of preceding correct POs when the current PO was also correct. This influence of current PO on preceding PO effects was shown to be dependent upon the probability of the preceding stimulus by the significant third-order interaction, Preceding Probability × Preceding PO × Current PO, $F$ (1, 19) = 4.37, $p < .05$. That is, when the current PO was correct, preceding correct POs facilitated the current choice RT regardless of preceding stimulus probability, however, when the current PO was incorrect, preceding correct POs facilitated the current reaction when the preceding stimulus was M but inhibited the current response when the preceding stimulus was L.

To study the nature of the interactions involving current stimulus probability, a separate 2 (Preceding Probability) × 2 (Current PO) × 2 (Preceding PO) ANOVA was calculated for each stimulus. The results of the ANOVA for Stimulus M indicated significant main effects of current PO, $F$ (1, 19) = 40.26, $p < .001$, and preceding PO, $F$ (1, 19) = 13.57, $p < .005$. The Preceding PO × Current PO interaction was significant, $F$ (1, 19) = 11.97, $p < .005$, as was the Preceding PO × Current PO × Preceding Probability interaction, $F$ (1, 19) = 7.76, $p < .025$. Figure 1 illustrates that for Stimulus M, the effect of preceding PO was influenced by both current PO and preceding probability as follows: Given a correct current PO, preceding correct POs shortened *S*s' choice reaction, but when the current PO was incorrect a correct preceding PO shortened *S*s' reactions when the preceding stimulus was M but lengthened *S*s' latencies when the preceding stimulus was L.

The ANOVA for the eight RT categories of Panel 2 in Fig. 1 indicated main effects similar to those found for Stimulus M: the effect of current PO, $F$ (1, 19) = 31.73, $p < .001$, was influenced by preceding probability; and the main effect of preceding PO, $F$ (1, 19) = 4.89, $p < .05$, was dependent on the current PO. That is, the Current PO × Preceding Probability interaction was significant, $F$ (1, 19) = 15.69, $p < .001$, as was the Current PO × Preceding PO, $F$ (1, 19) = 10.27, $p < .005$. The Current PO × Preceding Probability interaction demonstrated that when the current stimulus was L the effects of current PO were enhanced when the preceding stimulus was also L (i.e., a stimulus repetition). The Current PO × Preceding PO interaction was reliable since the effect of preceding PO was more prominent when the current PO was correct. Consequently, these separate analyses indicated a third-order interaction between preceding probability, preceding PO, and current PO in Panel 1 but not in Panel 2.

Two separate ANOVAs were performed to study the effects of preceding PO, current probability, and preceding probability for each of the two levels of current PO (i.e., solid lines vs. dotted lines in Fig. 1). Given a currently correct PO, *S*s' choice reactions were faster when the current stimulus was M, and regardless of the probabilities of the current or preceding stimuli *S*s' latencies were shortened by preceding correct POs. The ANOVA for the eight categories of correct PO indicated significant main effects of current probability, $F$ (1, 19) = 10.62, $p < .005$, and of preceding PO, $F$ (1, 19) = 40.27, $p < .001$. There was no reliable effect of preceding probability, and none of the interactions reached significance at the .05 level.

Similarly, within the current incorrect PO categories (dotted lines in Fig. 1), *S*s' choice reactions were faster when the current stimulus was M. Of the main effects, only that of current probability was significant, $F$ (1, 19) = 36.47, $p < .001$. In contrast to the observations for CP events, *S*s' reactions to an L were markedly slower when the preceding stimulus was also L, but *S*s' reactions to an M were somewhat faster when the preceding stimulus was L. Thus, the Current Probability × Preceding Probability interaction was reliable, $F$ (1, 19) = 6.72, $p < .025$. The Preceding PO × Preceding Probability Term, $F$ (1, 19) = 5.96, $p < .05$, was significant since, independently of current probability, *S*s' choice reactions were faster when preceded by CP rather than IP M but were slower when preceded by CP rather than IP L.

*Stimulus repetition effects.*—To study the effects of stimulus repetitions on choice RT, a 2 (Current Probability) × 2 (Alternation vs. Repetition, A/R, of the Stimulus) × 2 (Current PO) × 2 (Preceding PO) ANOVA was performed. A repetition was the second of any two consecutive identical stimuli; an alternation was the second of any two consecutive dissimilar stimuli. There was no main effect of the A/R variable ($F < 1$), but it was involved in three significant interactions, one being the fourth-order term: Current Probability × A/R × Current PO × Preceding PO, $F$ (1, 19) = 4.58, $p < .05$. As Fig. 1 indicates, this result was partly due to the marked slower reactions following a repetition of L (see the dotted lines in Panel 2 of Fig. 1). Because each of the significant interactions involved the current PO variable, a separate 2 (Current Probability) × 2 (A/R) × 2 (Preceding PO) ANOVA was performed for the CP and IP categories. For CP events, a significant effect of A/R was not found, nor were there any significant interactions. On the other hand, for IP events the A/R effect was reliable, $F$ (1, 19) = 6.88, $p < .025$, and of the interactions only Current Probability × A/R × Preceding PO was significant, $F$ (1, 19) = 6.07, $p < .025$. Like the fourth-order interaction described above, this interaction was largely due to the marked latency increase following a repetition of L.

*Order of sequential prediction outcome effects.*—In this analysis, 16 categories were defined by the POs on four consecutive trials. That is, *S*s were either correct or incorrect on each of four trials preceding a given choice reaction: the current trial, and the first, second, and third preceding trials. For example, letting C and I represent correct and incorrect POs, respectively, the fourth-order CCCI category would be the mean of every choice RT following an IP stimulus and preceded by

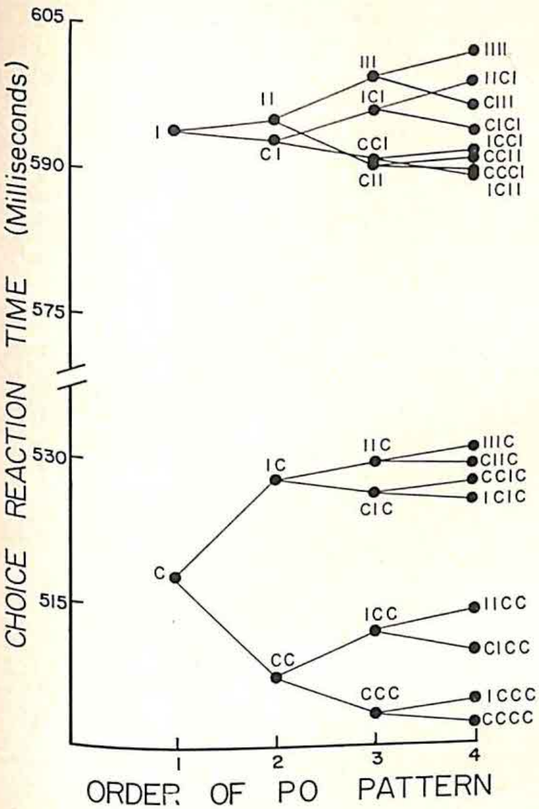


FIG. 2. Fourth-order tree diagram of correct and incorrect prediction outcomes.

at least three consecutive correct POs. As shown in the tree diagram of Fig. 2, there were marked second- and third-order effects of PO when the current PO was correct: choice reactions to second-order patterns were faster when the PO preceding that pattern was also correct. For incorrect current POs, choice reactions were faster when the preceding PO was correct, but the third-order categories were not in any consistent order.

The overall ANOVA, a 2 (Current PO) × 2 (First Preceding PO) × 2 (Second Preceding PO) × 2 (Third Preceding PO) design, was calculated with the 16 fourth-order categories of Fig. 2. The results indicated significant effects of current PO, $F\ (1, 19) = 40.87, p < .001$, first preceding PO, $F\ (1, 19) = 12.48, p < .005$, and second preceding PO, $F\ (1, 19) = 16.78, p < .001$, but not of the third preceding PO, $F\ (1, 19) = 2.22, .25 > p > .10$. Thus, a significant third- but not a fourth-order sequential effect of PO was demonstrated. Because of the difference in the magnitude of the first preceding PO effect between the current CP and IP categories, the Current PO × First Preceding PO interaction was significant, $F\ (1, 19) = 9.03, p < .001$; and, because there was an effect of second preceding PO and not of the third preceding PO, the interaction between these two variables was also reliable, $F\ (1, 19) = 4.87, p < .05$. None of the remaining interactions were significant, all $F$s $< 1$.

## DISCUSSION

The findings that *S*s reacted faster to the more probable stimulus alternative and to correctly predicted alternatives were consistent with recent choice RT studies which examined independent effects of probability and PO (e.g., Geller et al., 1971; Hinrichs & Craft, 1971). However, the reliable Stimulus Probability × PO interaction demonstrated herein was consistent with the results of Hinrichs and Craft but in contrast to those of Geller et al. That is, both Hinrichs and Craft and the present study found that the magnitude of current PO effects on choice RT was inversely proportional to the probability of the current stimulus, whereas the Geller et al. study found no such interaction. Consistent with the previous research by Whitman and Geller (1971, 1972), the present study found reliable second-order effects of PO. Moreover, the present research extended the sequential analysis of PO and demonstrated significant third-order effects.

The results of the present study indicated the necessity to consider both stimulus probability and PO in the analysis of stimulus repetition effects. Choice reactions to IP occurrences of the less probable stimulus were *slower* to a repetition than to an alternation, but reactions to CP occurrences of the less probable stimulus were *faster* to a repetition than to an alternation. Choice RT to the more probable stimulus was not significantly influenced by the alternation/repetition variable. Thus, such a third-order interaction indicates the necessity to categorize events according to PO and probability before determining stimulus repetition effects. Hence, the two analyses used by Geller et al. (1971) to study stimulus repetition effects were actually inadequate to evaluate the relationship between choice RT and stimulus run length.

One analysis examined the effects of stimulus run length for CP and IP events without considering stimulus probability and found that the choice RTs were faster to stimulus repetitions regardless of the current PO. The other analysis did not control for PO and found that choice RT to the less probable stimulus was slower to repetitions than to alternations but that reactions to the more probable stimulus were faster to repetitions than to alternations. However, in the present study, the effects of stimulus repetition were shown to interact with both PO and stimulus probability. Choice RT to repetitions of the less probable stimulus was more sensitive to current PO than was choice RT to any other second-order stimulus pattern.

Of interest are the implications which the effects of preceding PO and preceding probability have for expectancy notions which assume that choice RT to identify a particular stimulus varies inversely with *S*s' expectancy for the presentation of that stimulus (cf. Geller & Pitz, 1970). Because the effects of preceding PO and preceding probability were not greatly influenced by the current probability but were markedly different for CP and IP events, theoretical implications for expectancy theory will be considered separately for correct and incorrect predictions. Considering current correct predictions, there was no significant effect of preceding probability and no reliable interactions. Regardless of the probability of the preceding stimulus, a correct PO on the preceding trial facilitated choice RT relative to a preceding incorrect PO. To explain similar findings, Whitman and Geller (1971, 1972) suggested that increased confidence due to a correct PO augments *S*s' expectancy to perceive the predicted stimulus and consequently facilitates reactions to CP events and inhibits reactions to IP stimuli.

Choice RT to IP stimuli was influenced not only by the PO on the preceding trial but also by the probability of the preceding stimulus. That is, regardless of the current probability, correctly predicting the less probable stimulus on the preceding trial *inhibited* *S*s' reactions relative to those reactions preceded by incorrect predictions of the less probable stimulus, but correctly predicting the more probable stimulus on the preceding trial *facilitated* *S*s' reactions relative to those reactions preceded by incorrect predictions of the more probable stimulus. Because correctly predicting the less probable stimulus occurred on less than 10% of the trials, a reasonable post hoc explanation of the stimulus probability and PO effects found in the present study may be based on the assumption that correctly predicting the less probable stimulus causes a greater increase in *S*s' confidence in the next prediction than correctly predicting the more probable stimulus. Thus a preceding correct prediction of the less probable stimulus should cause a greater inhibition to respond to an incorrectly predicted stimulus than a preceding correct prediction of the more probable stimulus. Consequently, the findings that current PO had a greater influence on reactions to the less probable stimulus than on reactions to the more probable stimulus and that preceding probability had an effect on choice RT to nonpredicted events suggest that the effects of certain expectancy indicators (i.e., current PO and preceding probability) are most prominent when choice RT reflects inhibition by other expectancy indicators (i.e., current probability and current PO respectively).

## REFERENCES

GELLER, E. S., & Pitz, G. F. Effects of prediction, probability, and run length in discrete choice reaction speed. *Journal of Experimental Psychology*, 1970, **84**, 361–367.

GELLER, E. S., & WHITMAN, C. P. Probability, prediction, and incentive effects on choice reaction speed under two-dimensional stimulus presentations. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **6**, 67–68. (Summary)

GELLER, E. S., WHITMAN, C. P., WRENN, R. F., & SHIPLEY, W. G. Expectancy and discrete reaction time in a probability reversal design. *Journal of Experimental Psychology*, 1971, **90**, 113–119.

HINRICHS, J. V., & CRAFT, J. L. Verbal expectancy and probability in two-choice reaction time. *Journal of Experimental Psychology*, 1971, **88**, 367–371.

REMINGTON, R. J. Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, 1969, **82**, 250–257.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

WHITMAN, C. P., & GELLER, E. S. Runs of correct and incorrect predictions as determinants of choice reaction time. *Psychonomic Science*, 1971, **23**, 421–423.

WHITMAN, C. P., & GELLER, E. S. Prediction outcome, S–R compatibility, and choice reaction time. *Journal of Experimental Psychology*, 1971, **91**, 299–304.

(Received October 15, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 379–385

# THE USE OF CATEGORY INFORMATION IN A MEMORY-SEARCH TASK[1]

BARRY L. LIVELY[2] AND BARRY J. SANFORD

*Bowdoin College*

A memory-search experiment was performed in which size of the positive set and the relation between positive and negative sets were varied. Consonants and digits were present in the negative set, and either consonants or digits were present in the positive set. Thus, some stimuli in the negative set were of the same conceptual category as the positive set and others were of a different category. The slope of the RT function for the different category negative set stimuli was greater than zero but less than that of the function for negative set stimuli of the same category as the positive set. This result, among others, was discussed in terms of a hypothesis suggesting self-terminating search of the components (i.e., the conceptual category of the item and its name) of each member of the positive set but exhaustive search of the set.

In the memory-search task developed by Sternberg (1966, 1967), *S* is given a set of items to hold in memory. Individual items are then presented and *S*'s task is to indicate whether the presented item is a member of the memorized or positive set. Latency of response has been found to increase linearly with size of the positive set (Sternberg, 1966, 1967). Equal slopes of the reaction time (RT) functions for positive and negative responses suggest exhaustive serial scanning of the positive set independent of the nature of the response.

It is assumed (Sternberg, 1967) that the time between presentation of the item and the selection of a classification response is absorbed by at least two processing operations: (*a*) stimulus encoding, during which the presented stimulus is transformed into an internal representation, and (*b*) search, in which the internal representation resulting from the first operation is matched with representations of items in the positive set. A third operation might involve the selection and initiation of a response. The durations of these operations are assumed to be reflected in the function relating RT to the size of the positive set. The intercept consists of the combined times for the first and third operations, while the slope gives the average search time for one item.

The output of the stimulus encoding operation is some coded representation of the presented stimulus. The nature of that representation was investigated by Sternberg (1967). One hypothesis tested suggests that stimulus encoding produces an unprocessed direct copy of the presented stimulus. Another hypothesis holds that the name of the presented stimulus is produced. Blocks of trials were run in which the presented stimulus was degraded (i.e., covered with a checker-board pattern) or normal and size of the positive set was manipulated.

The direct copy hypothesis predicts that the time required for stimulus encoding would be the same for degraded and normal stimuli, but that search would take longer for a degraded stimulus since a "noisy" internal representation of the presented item is compared with the positive set. This implies equal intercepts but a greater slope for the degraded stimuli. The name hypothesis predicts more time for the name of the degraded stimulus to be extracted, but once available, the internal representation would be the same for degraded and normal stimuli; here, search time would be the same under the two conditions. This suggests a greater intercept for the degraded stimuli but equal slopes.

In the first of two sessions it was found zero slope for the ND RT function, since the sufficient information for responding is independent of size of the PS. For the NS stimuli, category information is not informative with respect to the response decision, and the members of the PS would have to be interrogated, yielding increasing RT with size of the PS. This hypothesis will be referred to as the category hypothesis.

---

[1] This research was supported by funds made available by the Bowdoin College Faculty Research Fund. The authors gratefully acknowledge the critical comments of Alfred H. Fuchs on an earlier draft.

[2] Requests for reprints should be sent to Barry L. Lively, Department of Psychology, Bowdoin College, Brunswick, Maine 04011.

Different predictions result if the encoded stimulus is a direct copy of the physical stimulus. Since there is no physical feature which reliably separates consonants from digits, the fact that a presented stimulus is drawn from the same or different conceptual category could have no bearing on search rate. This hypothesis (the direct copy hypothesis) would predict equal slopes for NS and ND RT functions.

## METHOD

*Subjects.*—The 24 *S*s were paid male undergraduate volunteers at Bowdoin College. The payoff scheme is given below.

*Design and procedure.*—Nature of the PS (consonants or digits) was manipulated as a between-*S* variable, while size of the PS was a within-*S* variable. Two groups of PSs were used for different *S*s in each category of positive set. The positive and negative sets used are given in Table 1.

Different sizes (s) of the PS were developed by starting with the item on the extreme left for a given group and adding successive items until the required number was obtained. The negative sets remained the same in all conditions of size of PS. The sequence of sizes of the PS was counterbalanced across *S*s. Within a block of trials, *S* was given 24 practice and 72 test trials. In both practice and test trials, the three classes of items appeared equally often. The order of kinds of trials was random except that no more than two stimuli of a given category appeared in succession.

The stimuli were typed on adding machine tape in elite type and were presented in the second field of a two-channel tachistoscope, the first field of which contained a random dot pattern. Presentation of the stimulus activated a timer which was stopped by a voice key. The verbal responses used were "BES" (for yes) and "BO" (for no); these responses were selected in an attempt to eliminate differential onset times for positive and negative responses.

At the beginning of a block of trials, *S* was shown a card containing the members of the PS. After memorizing the PS, the card was removed but made available between trials if requested.

The sequence on a given trial consisted of *E* saying "READY," followed 1–3 sec. later by the stimulus which remained visible until *S* had responded. The *S* was told whether he was correct or incorrect and was given his reaction time. After each block of trials, *S* was given a few minutes rest, followed by another block of trials with a different size of the PS.

TABLE 1

STIMULI USED IN POSITIVE AND NEGATIVE SETS FOR THE VARIOUS GROUPS

| Group | Positive set (PS) | Negative set same (NS) | Negative set different (ND) |
|---|---|---|---|
| 1 | 1467 | 2358 | BJLR |
| 2 | 2358 | 1467 | FHKN |
| 3 | BJLR | FHKN | 1467 |
| 4 | FHKN | BJLR | 2358 |

In order to maintain motivation, *S* was made aware of the following continuous payoff scheme. Pay equals

$$\$1.00 + \$.0025[1000 - \overline{\mathrm{RT}} - 5(\text{No. errors})],$$

where $\overline{\mathrm{RT}}$ is the mean correct RT. Thus, *S* was given \$1.00 base pay with the total amount earned dependent on speed and accuracy. The *S*s averaged about \$2.00 for the session.

## RESULTS

The dependent variables mean correct RT, fastest correct RT, and number of errors were computed for each condition for each *S*. While the RT treatment means and variances computed over *S*s were observed to covary and the treatment variances showed some tendency toward heterogeneity, transformation of the data was not required since the results of a Cochran's test for homogeneity of variance applied over all the RT treatment did not reach statistical significance, $g\ (k = 36, n = 11) = .065, p > .05$.

The major predictions concern mean RT for Cond. NS and ND. These data were subjected to a four-way analysis of variance, where order of testing PSs and nature of PS were treated as between-*S* variables and nature of the negative set and size of the PS were within-*S* variables.

Order of testing as a main effect did not significantly influence performance, $F\ (5, 12) = 1.28, p > .25$, although it produced a significant interaction with size of the PS, $F\ (10, 24) = 8.22,\ p < .001$.

that both the intercept and slope were higher for the degraded stimulus function than for the normal stimulus function; in the second session the slopes of the two functions were equal, but the intercept was greater for the degraded stimuli. From these results, Sternberg (1967) concluded that identification, or naming of the presented stimulus, need not precede classification of the stimulus. These results were confirmed by Chase (1969).

The Sternberg (1967) and Chase (1969) results suggested that the analysis of the stimulus in the first operation does not necessarily exceed a specification of the physical parameters of the stimulus whether in the form of a template or list of features. This interpretation has been pursued in investigations of the effects of visual and acoustic similarity on search rate. To the extent that the encoded stimulus shares visual or acoustic features with representations of the positive set, the search rate should be slowed. The direct copy hypothesis predicts a visual effect, while the name hypothesis predicts an acoustic effect. Experiments that have investigated these alternatives have produced mixed results. Congruent with Sternberg's conclusions, Chase and Posner (1965) found a slower search rate for visually confusable letters than for neutral control letters, but no difference between the control and acoustically confusable letters. In two experiments later performed by Chase and Calfee (1969), no effect of visual similarity was found, while acoustic similarity produced a retarded search rate in one experiment but not in the other. Thus, both visual and acoustic effects have been demonstrated, although the effect in neither case appears to be particularly reliable. It should be noted that an effect of acoustic similarity would follow from a name hypothesis only if the auditory or articulatory characteristics of the name are conceived as being part of the representation searched in memory. If a more abstract representation is assumed, one which has sensory inputs and can effect motor outputs but having neither auditory nor articulatory attributes itself, an effect of acoustic similarity would not be predicted. Thus the failure to uniformly find an effect of auditory similarity does not necessarily refute a name hypothesis.

Compatible with the notion of an abstract representation are the findings of DeRosa and Morin (1970). These investigators used sets of four consecutive digits as the positive set, with the negative set consisting of the remaining digits. It was found that as the difference between the arithmetic value of the presented stimulus and the highest (or lowest) digit in the positive set increased, RT decreased. This was true both for items in the negative set and for items in the positive set where reduced RTs were obtained for the middle stimuli of the positive set. Abstract properties of the stimulus, then, appear to be effective in determining search time.

The present experiment investigated another form of relationship between the presented item and the positive set. *S*s were given either consonants or digits in the positive set with both consonants and digits in the negative set. A presented item could be drawn from the positive set (PS), from the negative set of the same conceptual category (NS) as the PS, or from a negative set of a different conceptual category (ND). If the PS consisted of consonants, a negative set consonant would be designated NS and a negative set digit ND. Size of the PS was manipulated (one, two, and four items).

The question asked was whether the conceptual relationship between the presented item and the PS would influence search time. If the encoded form of the presented stimulus contains abstract information as suggested by DeRosa and Morin's (1970) results, it is possible that information about category membership of the presented stimulus is available. If this is the case, it is possible that ND stimuli would be rejected without exhaustive scanning of the PS. That is, if consonants are in the PS and a digit is presented, the mismatch between conceptual categories of the presented stimulus and the PS is sufficient to reject the digit as a member of the PS. This would suggest

Inspection of the means for $s = 1, 2$, and 4 as a function of number of prior blocks of trials revealed that $s = 4$ benefited more from practice than did the others. Size of the PS was found to be significant, $F (2, 24) = 74.0$, $p < .001$, while nature of the PS, i.e., whether it contained consonants or digits, did not produce a reliable difference, $F < 1$. In all further analyses, the data for consonants and digits are pooled.

With respect to testing the two hypotheses advanced, the critical tests concern the effect of the nature of the negative set and the interaction of size of the PS with nature of the negative set. The category hypothesis predicts zero slope for Cond. ND and a positive slope for Cond. NS, while the direct copy hypothesis predicts equal slopes for the two kinds of stimuli. The nature of the negative set, $F (1, 12) = 47.80$, $p < .001$, and its interaction with size of the PS, $F (2, 24) = 14.38$, $p < .001$, were both significant, arguing against the direct copy hypothesis. While the category hypothesis predicted this interaction, the zero slope predicted for Cond. ND was not obtained; a trend analysis performed on the ND data revealed a significant linear increase in RT as a function of size of the PS, $F (1, 46) = 39.12$, $p < .001$. The quadratic trend component was not significant, $F (1, 46) = 1.68$, $p > .10$. These findings argue against the category hypothesis.

Separate trend analyses were performed on the PS and NS RT functions. In both cases, only the linear component was significant, for PS, $F (1, 46) = 49.55$, $p < .001$, for the linear component, and $F (1, 46) = 1.73$, $p > .10$, for the quadratic component; for NS, $F (1, 46) = 59.05$, $p < .001$, for the linear component, and, $F (1, 46) = 1.21$, $p > .10$, for the quadratic component.

Since a significant linear component was obtained in all cases, best-fitting straight lines were estimated by the method of least squares. The data and straight lines are plotted in Fig. 1. Also given are equations for the straight lines.

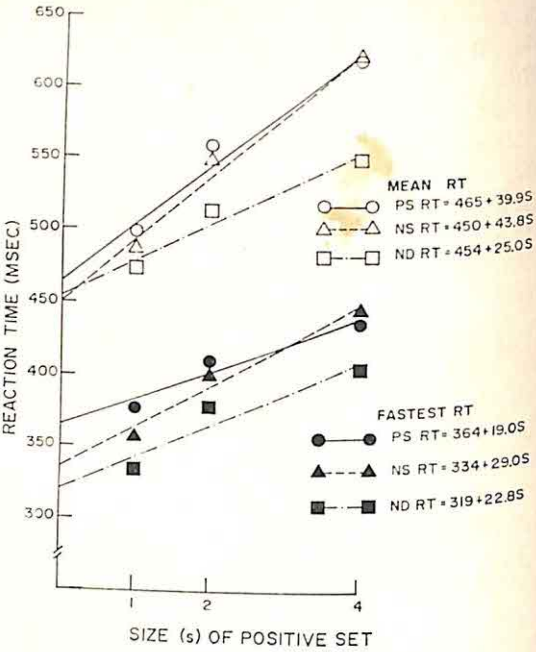


FIG. 1. Mean and fastest RT as a function of the size of the positive set.

It is seen in Fig. 1 that straight lines provide good fits to the obtained data, a result compatible with Sternberg's (1966, 1967) hypothesis of sequential search. Compatible with the notion of exhaustive search is the similarity of the slopes for the PS and NS conditions. An analysis of variance was performed with size of the PS and condition (PS vs. NS) as factors; neither condition ($F < 1$) nor the interaction between size of the PS and condition ($F < 1$) were significant. It also appears that the value of the intercept is not influenced by the relation between the presented stimulus and the PS. The values of the intercepts were computed for each *S* under each condition (PS, NS, and ND), and an analysis of variance performed. No difference was found among conditions ($F < 1$).

Before accepting the nonzero slope of the ND function as a genuine effect, contrary to both hypotheses advanced, the possibility must be considered that both hypotheses are in fact true and that the obtained ND function is a mixture generated either by some *S*s following the direct copy hypothesis and some the category hypothesis or by all *S*s responding according to one hypothesis on some trials and the other hypothesis on the remaining trials. Mixtures formed in either way would produce the obtained effect.

The first conjecture predicts that those *S*s following the direct copy hypothesis will evidence increasing RTs for ND stimuli as the number of items in the PS increases. Those *S*s following the category hypothesis would show no systematic change in RT to ND stimuli as size of the PS varies. Since for every *S* the RT for $s = 4$ was longer than that for $s = 1$, the between-*S* mixture hypothesis can be rejected.

The second conjecture was tested by extracting from each *S*'s data the fastest correct RT for each size of the PS under Cond. ND and averaging across *S*s. If the ND function for mean RT is a mixture of comparatively fast RTs generated under one hypothesis and slow RTs produced under the other, the fastest obtained RT ought to have occurred when *S* was following the category hypothesis, since this hypothesis, at least for the larger sizes of the PS, predicts faster RTs than the direct copy hypothesis. Furthermore, the slope of the RT function should be zero. A one-factor analysis of variance indicated a significant linear increase in RT with size of the PS, $F\ (1, 46) = 8.63$, $p < .01$. The quadratic component was not significant, $F\ (1, 46) = 1.94$, $p > .10$). The data along with the best-fitting straight line and its equation are given in Fig. 1. It is seen that the slope constants for ND mean and fastest RT functions are essentially the same. Evidently, performance in terms of the fastest RT is much like that of mean performance for Cond. ND. This conclusion is supported by the results of an analysis of variance comparing mean and fastest RTs for Cond. ND as a function of size of the PS in that size of the PS and mean versus fastest RT did not interact ($F < 1$). Thus the within-*S* mixture conjecture can also be rejected.

Analyses of fastest RT data are seldom if ever reported and, in view of the sensitivity of the ND fastest RTs to size of the PS, the characteristics of this dependent variable were further explored by obtaining for each *S* the fastest RT at each size of the PS for Cond. PS and NS. Separate one-factor analyses of variance indicated significant linear increases in RT for both Cond. PS, $F\ (1, 46) = 19.67$, and Cond. NS, $F\ (1, 46) = 31.72$, both $p\text{s} < .001$, as size of the PS increased. In neither case was the quadratic component significant, for Cond. PS, $F\ (1, 46) = 1.46$, $p > .10$; for Cond. NS, $F\ (1, 46) = 1.09$, $p > .25$. The data along with the best-fitting straight lines and equations are given in Fig. 1. The slopes of the fastest RT functions for Cond. PS and NS are seen to be lower than their counterparts based on mean performance. Analyses of variance show this to be a reliable reduction by demonstrating significant interactions between size of the PS and mean versus fastest RT for both Cond. PS, $F\ (2, 24) = 21.28$, $p < .001$, and Cond. NS, $F\ (2, 24) = 4.39$, $p < .025$. Further examination of the slopes of the fastest RT functions suggests that where for mean performance the slopes for the ND functions were lower than for Cond. PS and NS, for the fastest RT the slopes are all about the same. Analyses of variance comparing fastest RTs for Cond. ND separately with Cond. PS and with Cond. NS reveal no interaction between size of the PS and condition ($F < 1$ in both comparisons). Also, no interaction is evident in a similar comparison for Cond. PS and NS, $F\ (2, 21) = 2.27$, $p > .10$. These analyses also show that even for the fastest RT data latencies are lower for Cond. ND than for either Cond. PS, $F\ (1, 12) = 22.35$, $p < .001$, or Cond. NS, $F\ (1, 12) = 11.49$, $p < .01$.

Error data were also examined. The proportions of error responses are listed in Table 2 with the data for consonants and digits combined. The data are reported separately for false positives (saying yes to a negative set stimulus) and false negatives (saying no to a PS item). In the false positive data, it appears that errors increase as a function of size of the PS for Cond. NS, while remaining constant or even decreasing slightly for Cond. ND. Analysis of variance confirmed these observations, for Cond. NS, $F\ (2, 44) = 6.81$,

TABLE 2

PROBABILITY OF FALSE POSITIVE AND FALSE NEGATIVE RESPONSES AS A FUNCTION OF SIZE OF THE POSITIVE SET AND RELATION BETWEEN POSITIVE AND NEGATIVE SETS

| Type of error | S | | |
|---|---|---|---|
| | 1 | 2 | 4 |
| False positive NS | .023 | .038 | .080 |
| False positive ND | .024 | .019 | .017 |
| False negatives | .087 | .111 | .073 |

Note.—NS = a negative set of the same conceptual category as the positive set, ND, a negative set of a different conceptual category. S = set size.

$p < .01$, and Cond. ND, $F < 1$. Variation in false negatives as a function of size of the PS was not significant ($F < 1$).

## DISCUSSION

Neither the direct copy nor category hypothesis as stated earlier can account for the present data. The direct copy hypothesis can be rejected in view of the difference in slope between Cond. NS and ND. The category hypothesis fails in that the ND function has a positive slope.

The results indicate, as Sternberg (1967) had noted, that the characteristics of the stimulus involved in the search stage are more abstract than physical features. The present results argue that *S* can take advantage of the conceptual category of the stimulus, although perhaps the way in which the category information is used is not as simple as specified in the original hypothesis. Apparently, a more complicated process is involved than conditionalizing search on the basis of the category of the presented item.

How is the difference in slope between the NS and ND functions to be explained? It is assumed that the locus of the effect is in the search stage; the duration of this stage, unlike the stimulus registration and response selection stages, is assumed to vary with the number of items in the PS. The durations of presearch and postsearch stages are assumed constant and appear as the intercept of the RT functions. Since the intercepts of the RT functions for Cond. PS, NS, and ND were found to be the same in this experiment, the nonsearch stages are ruled out as loci of the effect.

The positive and equal slopes found for mean and fastest RT for Cond. ND suggest exhaustive scanning of the PS but as a faster rate than the mean search rate for NS items.

An explanation that predicts the major findings of this experiment might describe the memorial encoding of the PS to consist of each member of the PS represented as its name and conceptual category; thus, 1467 is coded as (DIGIT 1, DIGIT 4, DIGIT 6, DIGIT 7). Search of the set of "stimulus complexes" (i.e., category plus name) is assumed to be exhaustive across complexes but self-terminating within each complex with the order in which information is available about component matches or mismatches variable. For the examination of a single complex, acceptance or rejection at the level of the name component would constitute sufficient information for a decision, whereas rejection but not acceptance is possible at the level of the category component. For convenience, it is assumed that component matching and mismatching within a complex take equal amounts of time with the initiation of positive and negative overt responses also assumed equal. Total search time for a given presented stimulus would then consist of the sum of the examination times for the complexes in the PS. With 1467 as the PS, consider outcomes of presenting "B," "8," and "6"; ND, NS, and PS stimuli, respectively. The stimulus B differs from each member of the PS both in name and category, and as soon as information about one or the other is available a given complex could be rejected; the mean search time for 8 would consist of relatively fast times for trials on which the mismatch of the name was available first for most or all of the complexes and slower times for trials on which it was available second for most of the complexes. The outcome for a PS item would differ from that for an NS item only in that a match would be produced in one complex and a positive response would result. The component times would be the same. Thus, on the average, ND stimuli are processed faster than NS and PS stimuli.

In terms of mean RT, this component-search hypothesis predicts equal slopes for Cond. NS and PS and a lower but positive slope for Cond. ND. For the fastest RT, equal slopes are predicted in the limit for all conditions since presumably the fastest RTs under Cond. NS and PS are made when information about the name match is available first for most or all the complexes in the PS. All these predictions are upheld. One result not predicted by the hypothesis is that RT for Cond. ND is lower even for the fastest RT where slopes are essentially equal. This may indicate a check on accuracy of responding for the conceptually similar PS and NS stimuli.

The component-search hypothesis retains all the major features of the position adopted by Sternberg (1969) and extends that position only by articulating a mechanism in the search stage that takes advantage of learned characteristics of the presented stimulus. Some such modification seems necessary in view of the present results.

## REFERENCES

Chase, W. G. Parameters of visual and memory search. Unpublished doctoral dissertation, University of Wisconsin, 1969.

Chase, W. G., & Calfee, R. C. Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

Chase, W. G., & Posner, M. I. The effect of auditory and visual confusability on visual and memory search tasks. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1965.

DeRosa, D. V., & Morin, R. E. Recognition reaction time for digits in consecutive and nonconsecutive memorized sets. *Journal of Experimental Psychology*, 1970, **83**, 472–479.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

(Received October 18, 1971)

---

## ERRATA

In the monograph "A Multidimensional Scaling Study of Semantic Distance" by Jack B. Arnold in the October 1971 issue, the algebraic signs of the two terms forming the right-hand member of Equation 2 on page 354 should be changed, because as printed they define $-\log(m_j)$ instead of $\log(m_j)$. The clause immediately proceeding the equation should be corrected accordingly.

The right-hand member of Equation 3 on page 354 should appear under the radical as a square root.

Also, in Table 10, page 365, the column headings 1, 2, and 3 should be corrected to read 2, 3, and 4.

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 386–391

# EFFECTS OF MODALITY PREFERENCES ON PERFORMANCE ON A BISENSORY MISSING–UNITS TASK[1]

GARY M. INGERSOLL[2]
*Indiana University*

FRANCIS J. DI VESTA
*Pennsylvania State University*

The present investigation identified aural and visual modality preferences by performance on a bisensory digit-span task. Preference was defined by the magnitude of the difference in recall of stimuli presented via each of the modalities. Performance of aural and visual attenders was then compared on a second bisensory task; a modification of a missing-unit paradigm was used in which *S* was required to respond with missing units from two independent sets of stimuli, each set being presented on a separate channel. The results showed a disordinal aptitude treatment interaction between modality preferences and presentation modalities; i.e., aural attenders recalled more auditory than visual stimuli while visual attenders recalled more visual than auditory stimuli. In addition, differential recall patterns were found for each of the groups.

Research in the area of auditory and visual presentation modalities and their effects on learning and recall have generally been directed at a comparison of the efficiency of the two modalities. For example, Schulz (1969) has carried out a program of research designed to compare the relative effectiveness of visual and auditory presentation on various verbal learning tasks. Schulz, however, was forced to conclude that, by itself, presentation modality was not an overly potent variable. If modality effects do exist, he concluded, they must occur in the associative stage of learning.

Several investigators have studied the effects of presentation modality on short-term recall (e.g., Conrad & Hull, 1968; Craik, 1969), among which Murdock's (Murdock, 1966; Murdock & Walker, 1969) studies represent the most concerted attack on the role of auditory and visual presentation modalities in short-term memory. In an early investigation (Murdock, 1966), Murdock used a serial probe paradigm and demonstrated overall superior recall for auditory stimuli. Further, recall as a function of auditory stimuli displayed markedly greater recency effects, while recall of visual stimuli provided slightly greater primacy effects. However, due to the nature of the serial probe, primacy effects would be truncated since there is no probe for the first item in a list. Murdock and Walker used a free recall task and found essentially the same pattern with more clearly defined primacy effects for visual presentation. However, when Craik restricted recall order by instruction to start recall with the first items of the list rather than the last, he found strong primacy effects for both auditory and visual presentations. When recall was directed to begin with the last items in the list, the more traditional recency effects were found.

A task which has not been employed in the study of modality effects on recall, but which appears to have potential value for studying bisensory memory, is the missing-unit paradigm. Basically, this design consists of an initial definition of a finite set of units. All but one of the set is repeated in a random order and *S* is required to respond with the missing unit (Yntema & Trask, 1963). If, over several trials, words were systematically missing from each of the serial positions, a distribution of response probabilities should be found as a function of serial position. Such a distribution might offer some clarification of the nature of primacy effects for visual presentation.

---

[1] This article is based on a dissertation submitted by the first author in partial fulfillment of the requirements for the PhD degree at Pennsylvania State University. This study was partially supported by the Advanced Research Projects Agency (ARPA No. 1269), Office of Naval Research, Grant N00014-67-A-0385-0006, to the second author.

[2] Requests for reprints should be sent to Gary M. Ingersoll, Department of Educational Psychology, Indiana University, Bloomington, Indiana 47401.

Far less research has been conducted using bisensory auditory-visual presentation than research comparing the efficacy of either modality singly. Using a dichotic listening task, Broadbent (1954, 1956) presented digit spans simultaneously to each ear. He found that *S*s tended to recall the digit span from one ear and then the digit span from the other ear instead of recalling the digits in temporal pairs, as they were presented.

The aforegoing results were replicated using bisensory auditory-visual stimulation (Broadbent, 1956; Broadbent & Gregory, 1961). That is, when digit spans were presented bisensorily, *S*s tended to recall the set presented to one sense prior to the set presented to the other sense. Further, Dornbush (1968) found overall superiority of recall from auditory presentation when *S* is required to recall both sets from bisensory memory.

There are reports of only a limited number of studies in which individual differences have been studied under conditions of bisensory stimulation. This state of affairs is not overly surprising since the overall amount of bisensory research has not been sufficiently large enough to encourage the inclusion of individual difference variables. Broadbent (1956), however, observed that during recall, "there was no tendency for either sense to be universally the one delayed. In general, each subject has his own consistent preference for giving the eye or the ear first. [pp. 147–148]." This observation would be in accord with the limited capacity processing model which Broadbent proposes. However, the assumption that the first emitted response necessarily implies which modality was processed first may not be valid. The measure of response pattern is relatively unstable under different conditions and is susceptible to perceptual set (Senf, Rollins, & Madsen, 1967). Further, it is equally plausible that *S* could adopt a strategy of rehearsing the favored channel while quickly spewing out the information presented on the nonpreferred channel.

Madsen, Rollins, and Senf (1970) have studied the stability of response patterns to bisensory presentation across different presentation rates. It has been argued that as presentation rate decreased (pairs/sec) the ability to recall the information by pairs, as opposed to recall by modality, increased. Madsen et al. (1970) argued, however, that when *S*s were used as their own control in a within-*S*s design, their consistency across conditions was much more striking than any tendency to change with treatment conditions. Nevertheless, Madsen et al. did not compare individuals who consistently responded by modality to those who consistently responded by pairs on any other task.

Senf (1969) compared patterns of response by children with learning disorders (LDC) with patterns of response by normal control children (NCC) and supported the notion that the LDC group was less likely to construct intersensory associations than its NCC counterpart. Senf argued that such inability on the part of the LDC would restrict their development of reading skill. Within this framework, it was noted that the LDC group displayed a marked tendency to respond with the auditory channel first; whereas the NCC were fairly evenly distributed in the tendency to respond with either modality first. According to the limited capacity processing model, it could then be argued that the modality emitted first represented a preferred modality. However, Senf felt that response style was a source of confounding to the treatments in which he was most interested. Thus, he attempted to minimize the effect of response style by instructing *S*s to respond in a predetermined order.

The discovery of interactions between individual differences and treatments involving the two modalities would greatly enhance the study of presentation modalities, and their effects on learning and recall, by suggesting differential strategies employed by learners. The identification of differential patterns of recall as a consequence of these interactions for their respective modalities would clarify and provide further support for Murdock's

(1966) position. Accordingly, an important hypothesis in the present investigation was that visual attenders would recall more visual stimuli *and* exhibit marked *primacy* effects in recall. Conversely, aural attenders were hypothesized to recall more auditory stimuli *and* exhibit marked *recency* effects in recall.

## METHOD

A bisensory digit-span task, similar to that described by Broadbent (1956; Broadbent & Gregory, 1961) and Dornbush (1968) was used to determine whether *S* was an aural attender or a visual attender. Two digit spans were presented simultaneously; one digit span was presented visually and a second, different digit span was presented aurally. Modality preference was defined by *S*'s performance during the recall portion of five criterion trials. The *S* who tended to recall the visual digit spans more correctly than those presented aurally on those five criterion trials was classified as a visual attender. Conversely, the *S* who tended to respond more correctly to presentation via the auditory channel was classified as an aural attender.

A second task was a bisensory modification of the missing-unit paradigm described by Yntema and Trask (1963). In this procedure, five words were first presented to *S*. Then four of the five words were repeated in random order. The *S*'s task was to respond with the missing word. In the present investigation, two independent sets of five words were presented simultaneously, one visually and one aurally. Therefore, *S*'s task was to respond with two missing units, one from the set presented visually and one from the set presented aurally. Ten trials, in which words were dropped equally from each of the five serial positions, were administered.

An overall comparison of performance on the missing-units task as a function of auditory and visual presentation was made. However, in the principal analysis the performances of high visual attenders and high aural attenders were compared. The two extremes of modality preference served as a between-*S*s dimension. Thus, the design indicated a mixed analysis of variance with 2 between-*S*s dimensions (high visual attenders and high aural attenders) and 2 × 2 × 5 within-*S*s dimensions where the variables were responses to auditory and visual stimulation and 2 groups of 5 trials.

*Subjects.*—Undergraduate students at Pennsylvania State University who were enrolled in an introductory course in educational psychology served as *S*s. Participation was voluntary and not part of any course requirements. However, *S*s were awarded points toward their final grade for participating in the experiment. A total of 69 *S*s, all naïve to bisensory reception studies, participated.

*Stimuli, materials, and apparatus.*—The integers 1 through 0 (pronounced "zero") were used as stimuli for all digit spans. Integers were randomly assigned to four-digit digit spans with the restriction that the same figure could not occur twice within a bisensory trial. Thus, no number occurred twice within a single digit span nor did a given number ever occur in the digit span presented simultaneously on the other channel.

Stimuli for the missing-units task were randomly selected, without replacement, from a pool of 400 monosyllabic words found in the Thorndike-Lorge list of the 1,000 most common words. Homophones (e.g., son or sun) were eliminated from the pool of words to minimize ambiguity of the stimuli presented aurally. Proper nouns and numbers were also eliminated from the pool of words. Care was taken to insure that verbal associates did not occur within a set.

When presenting a *visual* stimulus array (e.g., a digit span) the members of the array can be presented at once (simultaneously) or the members of the array can be presented one at a time (sequentially). On the other hand, stimulus members of an *auditory* array can only occur sequentially. Thus, to maintain parallel modes of presentation, both auditory and visual stimulus arrays were presented sequentially.

Visual stimuli were rear projected onto a translucent screen, and auditory stimuli were presented via a tape recorder through an external speaker. The two signals were synchronized via a sound relay. Auditory stimuli on the magnetic tape were timed such that the "ready" signal initiated the presentation of the first slide of an array. As each slide appeared, an auditory signal initiated the next slide in the series. Presentation of stimuli on both modalities was suprathreshold and at a level not uncomfortable to *S*. The digits were presented at a rate of one every 2 sec. on each channel. The limitations of the apparatus restricted the exposure of the visual stimuli to approximately .8 sec. from onset to offset with a 1.2-sec. interstimulus interval. The duration of the auditory stimuli varied with a mean duration of approximately .65 sec.

*Procedure.*—The bisensory digit-span task and bisensory missing-units task were presented successively. Each *S* was instructed as to the nature of a bisensory digit-span task. He was instructed that for the first 10 bisensory trials, he would be told immediately prior to each trial which digit span (modality) he should recall first. The modality to be recalled first was randomly determined. Following these 10 trials, *S* was given 5 additional trials without preinstructions. He simply was told that he should recall the numbers in whatever order he felt was most comfortable for him. Performance on these 5 trials served as the measure of modality preference, which was defined as the magnitude of the difference between the number of auditory and visual stimuli recalled.[3] Responses were scored not only for correct retrieval of the units but also for retrieval of the correct temporal position; i.e., the inversion of two digits during recall would constitute an error.

---

[3] It may be that the term modality preference will eventually prove not to be totally accurate. How-

Upon completion of the bisensory digit-span task, *S* was given instructions describing the bisensory missing-units task. He was told that his task was to identify the missing word from two independent sets of information presented simultaneously over the two modalities. Written recall was used for both tasks.

## RESULTS

It was hypothesized that aural and visual preferences, as defined by the test measure, would interact with bisensory stimulation to affect differentially the recall of visual and auditory stimuli. It was also hypothesized that individuals who were characterized by one or the other modality preference would accentuate differences in those recall patterns typically associated with auditory and visual recall: that is, it was expected that the recall patterns of visual stimuli by visual attenders would be characterized by primacy effects, while the recall patterns of auditory stimuli by aural attenders would be characterized by recency effects.

Modality preferences were defined by the magnitude and direction of the differences between the auditory and visual scores on the test task. The distribution of scores was thus a distribution of difference scores, with negative scores indicating a visual bias and positive scores indicating an auditory bias. An analysis of the internal consistency of the difference scores yielded $r_{xx} = .65$. The scores of those *S*s who fell at the extremes of the continuum were selected for further analysis. For the initial study, an arbitrary cutoff point of ± 3 was established.

Earlier bisensory studies of short-term recall demonstrated that material presented by auditory means was found to be recalled more readily than units presented by visual means (Dornbush, 1968). Accordingly, an overall comparison of the mean number of correct responses for all *S*s to the auditory and visual missing units was made. This analysis yielded $t\,(68) = 1.50$, $p < .10$, one-tailed test. Examination of the distribution of responses as a function of presentation modality and the serial position from which the unit was eliminated failed to replicate ($p > .05$) the greater recency effects for recall of stimuli presented aurally or the primacy effects sometimes found for recall of stimuli presented visually when unisensory presentation procedures were employed (e.g., Murdock, 1966).

In the principal analysis of these data, however, performance of high aural attenders and high visual attenders, as defined by performance on the modality preference task, was compared. Of the 69 *S*s, 24 *S*s were identified as aural attenders and 16 *S*s were identified as visual attenders.

Neither the main effects of modality preference nor the mean number of items recalled as a function of presentation modality was significant ($p > .05$). However, the interaction of the effects of modality preference and presentation modality yielded, $F\,(1, 38) = 7.10$, $p < .05$. As hypothesized, the comparison of the means for visual attenders indicated that they recalled more visual ($\bar{X} = 3.06$) than auditory ($\bar{X} = 1.61$) missing words, $t\,(28) = 1.83$, $p < .05$, one-tailed test. Conversely, the aural attenders recalled more auditory ($\bar{X} = 3.08$) than visual ($\bar{X} = 1.79$ missing words, $t\,(38) = 2.02$, $p < .05$, one-tailed test.

The patterns of recall for the two groups across trials are displayed in Fig. 1a and 1b, which show that auditory recall by the aural attenders is markedly superior to recall of stimuli presented visually. The pattern is such that for aural attenders, the greatest difference in recall of auditory stimuli is found in the recent portion of the curve; however, while visual attenders clearly favor items presented visually over all trials, the largest differences between auditory and visual presentation are to be found in the earliest portion of the curve. A comparison of recall on the first serial position for visual ($\bar{X} = .44$) and aural ($\bar{X} = .22$) items by visual attenders yielded

---

ever, its use conforms to a limited number of references that are currently available. Eventually, it may be found that a label such as modality attenders, e.g., may be more accurate or descriptive.

preferences are stable and generalizable to more than one task. In addition, the performance of *S*s in the present experiment strongly suggests that preferences serve as monitoring or filtering systems which control the flow of information within the processing system.

Although the hypotheses for the present study were supported, replications with samples of *S*s from other populations, either different tasks, and a wider range of tasks, are clearly necessary. Certainly in the interest of establishing the generality of the present results, replications with a more heterogeneous population than college students are required. The use of a more heterogeneous group (e.g., elementary school children) would allow the incursion of a wide range of cognitive variables to affect performance. If the results are replicable with such populations, the construct of modality preferences could be presumed to be distributed over a wide range of *S* populations. The use of a more variable sample would also allow the selection of more extreme scores from the ends of the preference continuum. More extreme aural and visual attenders might produce response patterns more clearly associated with the predictions derived from a separate sensory storage model.

## REFERENCES

Broadbent, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191–196.

Broadbent, D. E. Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1956, **8**, 145–152.

Broadbent, D. E., & Gregory, M. On the recall of stimuli presented alternately to the sense organs. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 103–109.

Conrad, R., & Hull, A. J. Input modality and the serial position curve in short-term memory. *Psychonomic Science*, 1968, **10**, 135–136.

Craik, F. I. M. Modality effects in short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 658–664.

Dornbush, R. L. Input variables in bisensory memory. *Perception & Psychophysics*, 1968, **4**, 41–44.

Madsen, M. C., Rollins, H. A., & Senf, G. M. Variables affecting immediate memory for bisensory stimuli: Eye–ear analogue studies of dichotic listening. *Journal of Experimental Psychology*, 1970, **83** (3, Pt. 2).

Murdock, B. B., Jr. Visual and auditory stores in short-term memory. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 206–211.

Murdock, B. B., Jr., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

Schulz, R. W. *Learning of aurally received verbal material: Final report.* (Contract OE5-10-018) Iowa City: University of Iowa, March 1969.

Senf, G. Development of immediate memory for bisensory stimuli in normal children and children with learning disorders. *Developmental Psychology*, 1969, **1** (6, Pt. 2).

Senf, G. M., Rollins, H. A., & Madsen, M. C. Recall order of dual-channel input as a function of induced recall strategy and stimulus rate. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, **2**, 57–58. (Summary)

Yntema, D. B., & Trask, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

(Received October 18, 1971)

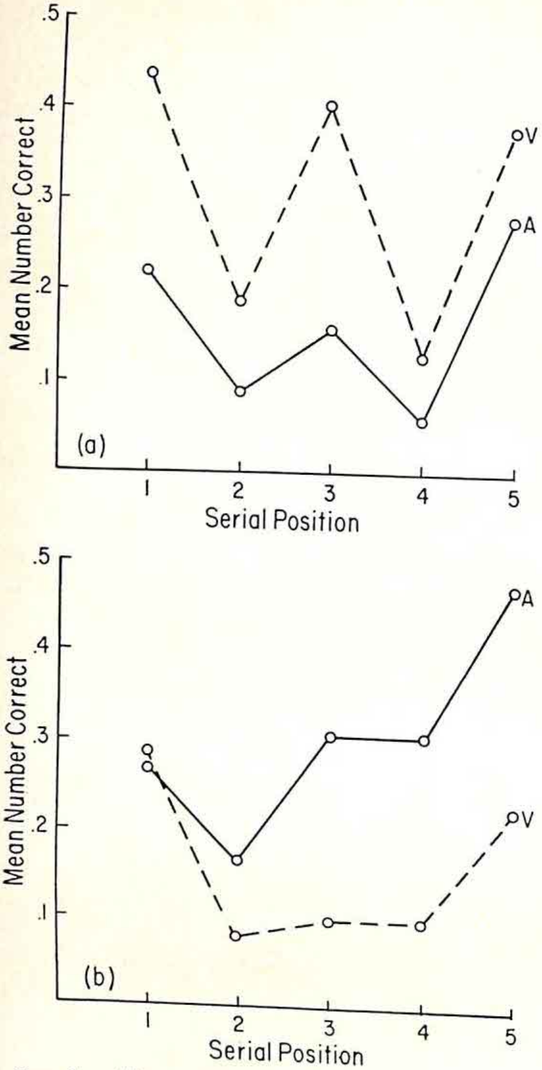


FIG. 1. (a): Auditory (A) and visual (V) recall during the missing-units task for visual attenders. ((b): Auditory (A) and visual (V) recall during the missing-units task for aural attenders.)

$t$ (158) = 2.24, while a comparison of recall on the last serial position for aural ($\bar{X}$ = .48) and visual ($\bar{X}$ = .23) items by aural attenders yielded $t$ (158) = 3.13, both $p$s < .05.

## DISCUSSION

The results of this experiment indicated that visual and aural modality preferences were clearly stable across two independent tasks involving short-term memory. In addition, the definitions of aural attenders and visual attenders were sufficiently distinct to warrant further investigation.

The results based on *S*s' performance on the missing-units task reaffirm findings from studies based on unisensory presentation of materials, whether auditorally or visually, (e.g., Murdock, 1966; Murdock & Walker, 1969) and from experiments based on bisensory, auditory-visual presentations with short-term memory (Dornbush, 1968). In these two groups of studies, as in the present study, the recall of auditory stimuli was superior to the recall of visual stimuli. The overall results also indicated greater recency effects for recall of stimuli presented aurally. However, a greater primary effect for recall of stimuli presented visually was not obtained.

The results of the analysis of data based on the performance of *S*s clearly identified as visual attenders and aural attenders negated the significant main effect of recall as a function of modality of presentation. Furthermore, the overall level of performance for the two groups did not differ; that is, both aural and visual attenders perform equally well on the missing-units task as a whole. Parenthetically, this does not preclude the possibility that the two groups may differ in performance on another type of task, especially one in which the presentation of information by one or the other of the modalities is favored. However, when performance on the missing-units task was analyzed as a function of modality preferences *and* presentation modalities, the hypothesized disordinal interaction was found. Visual attenders correctly identified more visual units than auditory units; conversely, aural attenders correctly identified more auditory than visual units.

The hypothesis was clearly supported that aural attenders and visual attenders would recall differentially stimuli presented aurally and visually. Further, the distribution of responses across serial positions corroborated the hypothesized accentuation of different recall patterns to stimuli presented visually and aurally by *S*s identified as having congruent modality preferences. Thus, visual attenders not only correctly identified more visually than aurally presented missing units, but the accuracy of their recall was greatest when the unit to be identified occupied initial serial positions (primacy). Aural attenders, on the other hand, not only correctly identified more stimuli presented aurally than visually, but the accuracy of their recall was greatest when the units to be identified had been eliminated from the latter part of the set (recency). These results imply that modality

Journal of Experimental Psychology
1972, Vol. 93, No. 2, 392-397

# INCENTIVE EFFECTS IN THE DISCRIMINATION OF TWO SIMULTANEOUSLY OCCURRING PROBABILITY DISTRIBUTIONS[1]

E. SCOTT GELLER,[2] CHARLES P. WHITMAN, AND GLENDA J. McGUIRE

*Virginia Polytechnic Institute and State University*

Incentive conditions were manipulated between and within *S*s in a two-dimensional probability learning experiment. Prior to stimulus presentations, *S*s predicted which of two symbols they expected and in which of two locations the symbol would appear. By independently varying the relative probabilities of the two symbols and of the two locations, the effects of different incentive conditions on *S*s' discrimination of two simultaneously occurring probability distributions were studied. The prediction distributions of a no-incentive group and a group receiving a reward only when correctly predicting both dimensions "matched" the theoretical probabilities of each dimension. When payoffs followed correct predictions of either dimension or of only one dimension, prediction proportions for the more frequent alternatives underapproximated the probabilities. When *S*s were exposed to two incentive levels, negative but not positive contrast occurred.

Several investigators have used a probability learning (PL) paradigm to study the effects of incentives on human choice behavior. The *S*s predict each occurrence of sequentially presented stimuli, and incentive effects are studied by varying the incentive magnitude for correct predictions between experimental groups or by exposing the same *S* to more than one incentive level. A majority of the experiments using a between-*S* design have shown no effects of incentive on PL (e.g., Halpern, Schwartz, & Chapman, 1968; Myers, Fort, Katz, & Suydam, 1963; Schnorr, Lipkin, & Myers, 1966). However, Siegel (1964), using college students or young children as *S*s, found that the prediction proportions of no-incentive groups "matched" the stimulus probabilities within 100 trials, while the prediction proportions of groups receiving a reward for each correct prediction (i.e., 5¢ for students, a trinket for children) overestimated the probability of the more frequently occurring stimulus within 100 trials. Conversely, Das and Panda (1963) found an interfering or suppressive effect of an incentive for correct predictions. Specifically, children who received no payoff for correct predictions "matched" the theoretical probabilities within 100 trials, whereas the prediction distributions of children who were given candy following correct predictions significantly underestimated the probability of the more frequently occurring stimulus.

Two basic methodologies have been used to study the effects of incentive magnitude on PL in a within-*S* design, i.e., successive and simultaneous paradigms. In the successive manipulation of incentive *S*s are given initially one level of incentive for several prediction trials and subsequently shifted to a higher or lower level. Those PL studies that incorporated this successive manipulation of incentive observed negative but not positive contrast (Halpern et al., 1968; Swensson, 1965). That is, after *S*s received a shift from a high to a low level of incentive, their prediction proportions for the more frequently occurring stimulus were lower than those of control *S*s who received the same low incentive magnitude throughout the experimental

---

[1] The authors thank J. J. Franchina for his helpful suggestions. For assistance in collecting data, we thank J. C. Farris, E. K. Grove, S. O'Quinn, D. S. Post, D. I. Shedivy, W. G. Shipley, E. L. Viers, and R. G. Wiley; for their donation of electronic supplies, we thank Western Electric Company, Inc. Portions of this research were presented at the meeting of the Midwestern Psychological Association, Detroit, May 1971.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virgina Polytechnic Institute and State University, Blacksburg, Virginia 24061.

session. However, when switched from low to high incentive levels, *S*s did not predict the more probable stimulus more often than *S*s receiving the same high incentive magnitude.

Negative but not positive contrast was also demonstrated when the payoff *S* received for a correct prediction was varied randomly on each trial (Halpern et al., 1968; Lipkin, 1968; Schnorr et al., 1966). For this within-*S* manipulation of incentive, one of two incentive magnitudes was stamped on the back of each card in a shuffled deck (e.g., 1¢ or 5¢), and when predicting which of two stimulus alternatives was on the face of a given card, *S*s "risked" the incentive magnitude indicated on the back. Although termed a simultaneous paradigm, with such a PL design only one incentive level is operating on any given trial, and therefore the distinction between the successive and simultaneous variations of incentive magnitude is based upon the number of times that the incentive levels are switched in a given experimental session. With the successive paradigm, the incentive level is changed once; in a simultaneous design, the incentive magnitude changes numerous times during a session.

In order to present *S*s with two incentive levels simultaneously on each prediction trial, the present experiment varied incentive conditions in a two-dimensional PL task. Specifically, stimulus presentations were differentiated into two dimensions: shape and location. The relative frequencies of two stimulus alternatives of each dimension varied according to independent probability schedules and, prior to each presentation, *S*s predicted the alternatives of each dimension. Consequently, when a payoff followed correct predictions of only one dimension, *S*s were exposed to two different incentive levels on each prediction trial (reward and no reward). Since such a design permits the incentive for correct predictions of a given dimension to remain constant throughout an experimental session, within-*S* incentive effects were studied independently of shift effects.

## METHOD

*Subjects.*—One hundred volunteers (65 males, 35 females) from the introductory psychology course at Virginia Polytechnic Institute and State University served as *S*s.

*Stimuli.*—On each trial, the symbol *∩* or *U* was presented by the illumination of one of two miniature readouts (5 × 1 cm.). The readouts were located behind a one-way mirror and were separated by a horizontal axis of 35 cm. During an experimental session, the room was dimly lit so that *S* only saw through the mirror screen when the readout was illuminated.

The relative probability of a *∩* or *U* symbol, verbally labeled "down" (D) and "up" (U), respectively, and a "right" (R) or "left" (L) location varied on independent 70-30 probability schedules. Specifically, a sequence of 200 stimulus presentations was randomly determined by filtering a uniform random-number generator on an IBM 360 computer system according to the following relative frequencies specified by the product of the dimension alternatives of each shape-location combination: DR (down-right) = .70 × .70 = .49; DL (down-left) = .70 × .30 = .21; UR (up-right) = .30 × .70 = .21; and UL (up-left) = .30 × .30 = .09.

*Incentives.*—Those *S*s who received an incentive for correct predictions were told that they were in a prediction contest, and that those students who earned the highest score in their group would receive 5 extra points added to his point total for the introductory psychology course. During the experimental session, *S*'s prediction score was cumulatively recorded on a contest counter located below the stimulus screen and directly in front of *S*. After a session, *S*'s final score was posted on a chart and *S* could then compare his score with the scores earned by other *S*s. The type of prediction outcome that increased *S*s' prediction scores was varied between experimental groups. Five groups of 13 males and 7 females each were used, each group receiving one of the following five conditions: (*a*) no contest counter (Group No Counter), (*b*) score on counter increased only following correct predictions of *both* the shape and location dimensions (Group Both), (*c*) score on counter increased following *either* a correct shape or location prediction (Group Either), (*d*) score on counter increased only following correct predictions of the *shape* dimension (Group Shape), (*e*) score on counter increased only following correct *location* predictions (Group Location).

*Procedure.*—The *S*s performed individually in 1-hr. sessions. At the start of each session, the two-dimensional prediction task was explained and 10 practice trials were administered during which questions concerning the task were answered. The *S*s were told that the stimulus sequence was preprogrammed on paper tape, but were given no information concerning the nature of the sequence. All *S*s were encouraged to make as many correct shape and location predictions as possible. The *S*s in the incentive conditions (i.e., Groups Both, Either, Shape, and Location) were told that they were participating in a prediction contest and were given appropriate instructions specifying which type of prediction outcome influenced their score on the contest counter.

For each of 200 trials, the order of events was as follows: (*a*) *S*'s two-dimensional stimulus prediction, (*b*) a 1-sec. "attention buzzer" indicating a subsequent stimulus presentation, and (*c*) a 3-sec. presentation of the two-dimensional stimulus. For the incentive conditions, the score on the contest counter was increased by 1 immediately following a stimulus presentation if the appropriate prediction had been correct.

## RESULTS

For each successive block of 20 trials, the proportion of *S*s in each group who predicted the most frequent alternative of each dimension, D and R, was calculated. The overall analysis of variance (ANOVA), a 5 (Groups) × 2 (Dimensions) × 10 (Trial Blocks), indicated significant main effects of group, $F\ (4,\ 95) = 5.40,\ p < .001$; dimension, $F\ (1,\ 95) = 9.96,\ p < .005$; and trial block, $F\ (9,\ 855) = 63.62,\ p < .001$. Of the interactions, the Group × Trial Block, $F\ (36,\ 855) = 1.81,\ p < .005$, and Dimension × Trial Block, $F\ (9,\ 855) = 3.31,\ p < .001$, were significant. To determine specific incentive effects, the predictions proportions for each incentive group were compared with those of the no-incentive condition using error estimates provided by the overall ANOVA.

*Incentive magnitude effects.*—Figure 1 depicts a comparison of the prediction proportions for the no-incentive condition (Group No Counter) with those of Group Both and Group Either. The prediction proportions of Groups No Counter and Both eventually matched the theoretical probabilities of each dimension alternative. Apparently, the incentive of a contest point for correctly predicting the alternatives of both dimensions did not influence *S*'s predictions relative to no-incentive controls. The ANOVA comparing Groups No Counter and Both, a 2 (Groups) × 2 (Dimensions) × 10 (Trial Blocks) factorial, indicated a reliable main effect of only trial block, $F\ (9,\ 855) = 44.19,\ p < .001$, and none of the interactions were significant, all $F$s $< 1$.

The comparison between the prediction proportions of Groups No Counter and Either in Fig. 1 indicates that the effect of an incentive for correct predictions of either dimension alternative was to depress the prediction proportions for D and R. Specifically, the proportion of D and R predictions for Group Either did not "match" the theoretical probabilities, but

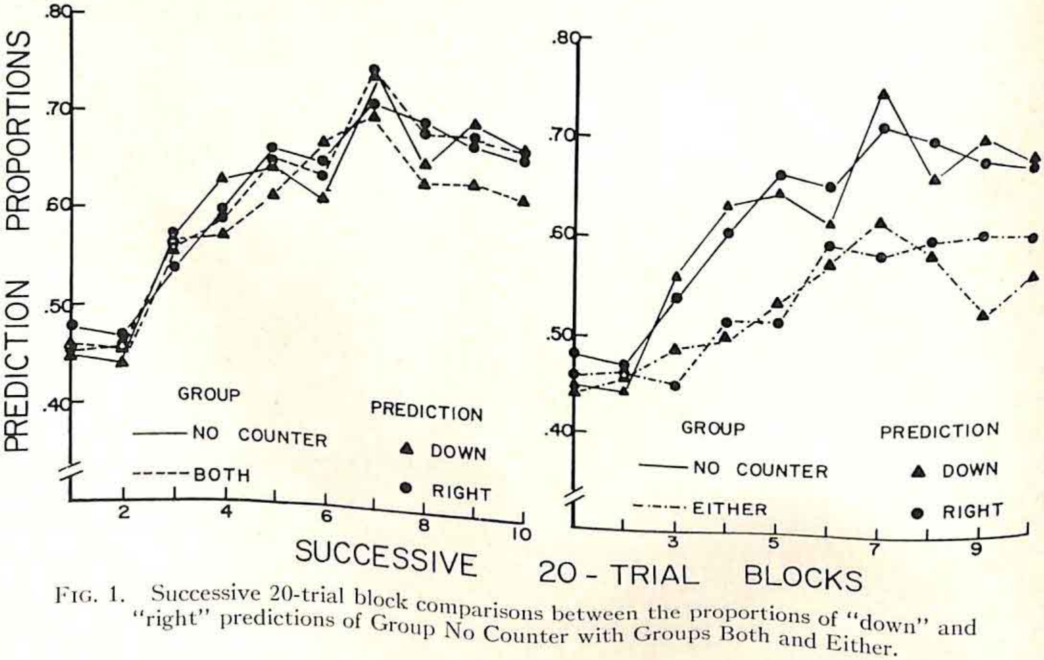


FIG. 1. Successive 20-trial block comparisons between the proportions of "down" and "right" predictions of Group No Counter with Groups Both and Either.

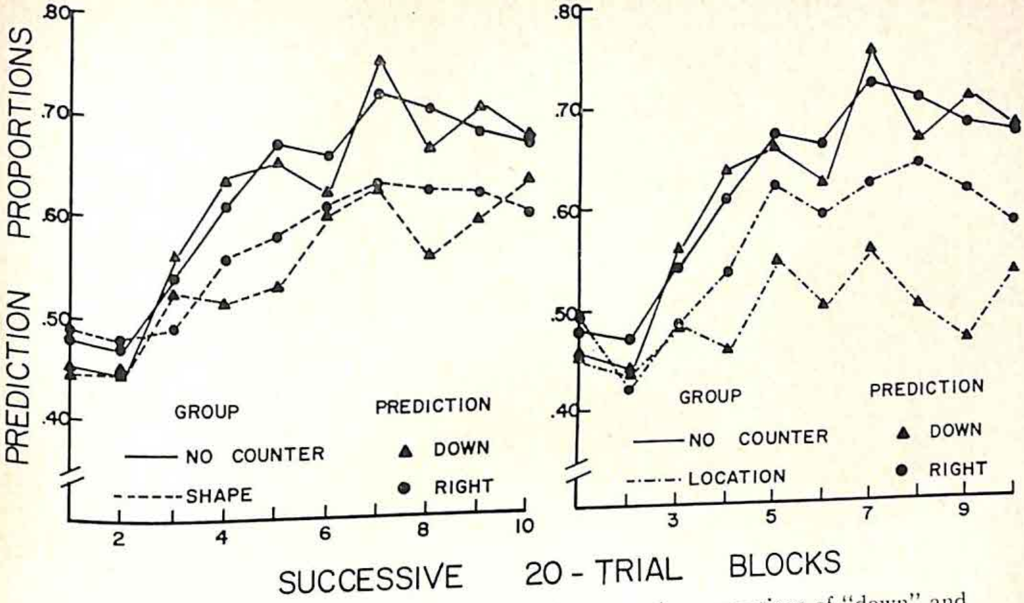


FIG. 2. Successive 20-trial block comparisons between the proportions of "down" and "right" predictions of Group No Counter with Groups Shape and Location.

after 100 trials reached an asymptote markedly below .70. The ANOVA indicated significant main effects for group, $F$ (1, 95) = 10.41, $p < .005$, and trial block, $F$ (9, 855) = 30.09, $p < .001$, but not for dimension, $F < 1$. Of the interactions, only the Group × Trial Block was reliable, $F$ (9, 855) = 3.14, $p < .001$.

The effects of an incentive for correct predictions of only the shape dimension or only the location dimension are illustrated in Fig. 2. The comparison between Groups No Counter and Shape is quite similar to the comparison in Fig. 1 between Groups No Counter and Either. That is, the effect of rewarding only correct predictions of the shape dimension was to depress the prediction proportions of both D and R below their theoretical probabilities. Correspondingly, the ANOVA indicated significant main effects for group, $F$ (1, 95) = 5.91, $p < .025$, and trial block, $F$ (9, 855) = 32.23, $p < .001$, but not for dimension, $F < 1$. Of the interactions, only the Group × Trial Block reached significance, $F$ (9, 855) = 2.25, $p < .025$.

The comparison between the prediction proportions of Groups No Counter and Location in Fig. 2 indicates that the incentive for correct location predictions not only lowered the prediction proportions of D and R, but within Group Location the proportion of D predictions was markedly below the proportions for R. The ANOVA indicated significant main effects of group, $F$ (1, 95) = 13.66, $p < .001$, dimension, $F$ (1, 95) = 7.11, $p < .01$, and trial block, $F$ (9, 855) = 28.02, $p < .001$. In addition, all interactions were significant, all $p$s $< .05$. The simple effects analysis showed that the triple interaction was due to a reliable difference between D and R for Group Location but not for Group No Counter. That is, the ANOVA comparing the prediction proportions of D and R for Group Location indicated a significant main effect of dimension, $F$ (1, 95) = 11.95, $p < .001$, and trial block, $F$ (9, 855) = 7.20, $p < .001$, and a significant Dimension × Trial Block interaction, $F$ (9, 855) = 3.89, $p < .001$.

*Incentive contrast effects.*—Negative contrast effects are indicated for Groups Location and Shape by comparing the prediction proportions for the nonrewarded dimension (i.e., location for Group Shape and shape for Group Location) with the prediction proportions of Group No Counter. Specifically, in Fig. 2 negative contrast is shown for Groups Shape and

Location by the lower proportion of R predictions for Group Shape than for Group No Counter, and the lower proportion of D predictions for Group Location than for Group No Counter. The ANOVA comparing the R proportions between Groups Shape and No Counter indicated significant main effects of group, $F$ (1, 95) = 3.52, $p < .05$, and trial block, $F$ (9, 855) = 15.16, $p < .001$; and no interaction, $F < 1$. Likewise, significant main effects of groups, $F$ (1, 95) = 11.53, $p < .001$, and trial block, $F$ (9, 855) = 11.18, $p < .001$, were obtained from the ANOVA comparing the D proportions between Groups Location and No Counter. The Group × Trial Block interaction was also significant, $F$ (9, 855) = 4.79, $p < .001$.

Comparing the prediction proportions for the rewarded dimensions of Groups Location and Shape with the appropriate prediction proportions of Group Both demonstrates that positive contrast did not occur. In fact, the graphical comparison between the R proportions of Group Both (Fig. 1) with the R proportions of Group Location (Fig. 2) suggests that the prediction proportions of R were actually lower when correct predictions of location were rewarded in the context of nonreward. The ANOVA that compared the proportion R predictions between Groups Both and Location indicated a significant main effect of group, $F$ (1, 95) = 4.01, $p < .05$, and trial block, $F$ (9, 855) = 19.24, $p < .001$, and no interaction, $F < 1$.

The ANOVA between the proportion of D predictions of Groups Shape and Both also indicated an absence of positive contrast. A main effect of trial block was obtained, $F$ (9, 855) = 13.95, $p < .001$, but neither the main effect of group nor the Group × Trial Block interaction were significant ($p$s > .10).

## DISCUSSION

By manipulating the type of prediction outcome that was followed by a payoff, the present research demonstrated the extent to which specific incentive conditions affected differential processing of shape and location dimensions. The comparison of probability learning (PL) between *S*s receiving an incentive for correctly predicting both dimension alternatives and no-incentive controls supported the conclusion of several others, that between-*S* incentive manipulations do not influence PL (e.g., Halpern, et al., 1968; Myers et al., 1963; Schnorr et al., 1966). On the other hand, the between-*S* comparisons between Group Counter and the other incentive groups (Groups Either, Shape, and Location) indicated less efficient probability discrimination when *S*s received an incentive, and consequently replicated the earlier findings of Das and Panda (1963) and Stevenson and Weir (1959).

The depressed prediction proportions for the more probable dimension alternatives of Groups Either, Shape, and Location might be explained by an interfering influence of a reward presentation. That is, rewards had no informative value in the present experiment and since the rewarding occurred as an independent event on each trial it could have distracted *S*s from developing an effective probability discrimination (cf. Das, 1969, p. 49). For example, when correct predictions of either dimension were followed by a reward three different types of prediction outcomes were reinforced (i.e., rewards followed correct predictions of both dimensions, only shape, and only location). Besides distracting *S*s from their discrimination task, such an incentive condition, which treated each dimension separately when determining payoffs, probably influenced *S*s to process each dimension individually rather than jointly, and consequently less efficiently. It is noteworthy that the incentive condition in which the stimulus dimensions were not considered apart when determining payoffs (i.e., Group Both) did not inhibit *S*s' discrimination performance.

Since rewarding correct predictions of only one dimension should influence *S*s to process each dimension individually rather than jointly, the underpredictions of the 70% alternatives by Groups Location and Shape could be explained with interference notions. However, additional post hoc speculations are necessary to account for the observation of a within-*S* effect of incentive level for Group Location but not for Group Shape. Specifically, *S*s in Groups Location and Shape were exposed to two incentive levels on each prediction trial, but only *S*'s of Group Location showed differential PL for the rewarded and nonrewarded dimensions. To account for these differences, the notion of a priority hierarchy for processing multiple stimulus dimensions might be considered (cf. Geller & Whitman, 1971). Since *S*s must locate a stimulus prior to categorizing its shape, it is reasonable to assume that *S*s consistently processed location before shape. Consequently, when rewarded for correct location predictions *S*s' attention to the probabilities of the shape alternatives was inhibited. However, given that *S*s must first attend to location prior to categorizing shape, reward for only correct shape predictions could not differentially influence the discrimination of the dimension probabilities. The recent findings of Geller and Whitman are supportive of these notions that the location of a stimulus is consistently categorized before shape, and that incentive effects of multiple-choice behavior are dependent upon whether one dimension receives selective attention. When stimulus presentations varied simultaneously on shape and location dimensions in a choice reaction time paradigm, Geller and Whitman found that *S*s responded markedly faster when identifying the location rather than the shape of each two-dimensional presentation. Moreover, the relative frequency of a location alternative significantly influenced *S*s' reaction time to identify the shape of a presentation; but the probability of a given shape did not influence *S*s' reaction time to identify a location alternative.

When *S*s experienced two incentive levels on each prediction trial, their prediction proportions for the more frequent alternative of the nonrewarded dimension fell reliably below those of *S*s in the no-incentive control group (i.e., negative contrast). Since these *S*s of Groups Shape and Location did not predict the more probable alternative of the rewarded dimension more often than did *S*s who were rewarded for correctly predicting both dimensions (Group Both), positive contrast was not demonstrated. Such observations of negative but not positive contrast are consistent with the findings of a majority of learning experiments involving successive or simultaneous shifts in reward magnitude. Specifically, investigators who varied incentive levels within *S*s in one-dimensional PL have typically demonstrated only negative contrast (e.g., Halpern et al., 1968; Lipkin, 1968; Schnorr et al., 1966). Additionally, incentive contrast studies with animals have consistently found negative contrast, and have only rarely demonstrated positive contrast (see review by Black, 1968).

## REFERENCES

Black, R. W. Shifts in magnitude of reward and contrast effects in instrumental and selective learning. *Psychological Review*, 1968, **75**, 114–126.

Das, J. P. *Verbal conditioning and behavior.* New York: Pergamon Press, 1969.

Das, J. P., & Panda, K. C. Two-choice learning in children and adolescents. *Journal of General Psychology*, 1963, **68**, 203–211.

Geller, E. S., & Whitman, C. P. Probability, prediction, and incentive effects on choice reaction speed under two-dimensional stimulus presentations. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **6**, 67–68. (Summary)

Halpern, J., Schwartz, J. A., & Chapman, R. Simultaneous and successive contrast effects in human probability learning. *Journal of Experimental Psychology*, 1968, **77**, 581–586.

Lipkin, S. G. An analysis of the interaction of incentive magnitude and relative event frequency in probability learning. *Canadian Journal of Psychology*, 1968, **22**, 1–10.

Myers, J. L., Fort, J. G., Katz, L., & Suydam, M. M. Differential monetary gains and losses and event probability in a two-choice situation. *Journal of Experimental Psychology*, 1963, **66**, 521–522.

Schnorr, J. A., Lipkin, S. G., & Myers, J. L. Level of risk in probability learning: Within- and between-subjects designs. *Journal of Experimental Psychology*, 1966, **72**, 497–500.

Siegel, S. *Choice, strategy, and utility.* New York: McGraw-Hill, 1964.

Stevenson, H. W., & Weir, M. W. Variables affecting children's performance in a probability learning task. *Journal of Experimental Psychology*, 1959, **57**, 403–412.

Swensson, R. A. Incentive shifts in a three-choice decision situation. *Psychonomic Science*, 1965, **2**, 101–102.

(Received October 18, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 398–403

# BLOCKED PRESENTATION IN MULTITRIAL FREE RECALL[1]

RICHARD M. WEIST[2]
*University of Nebraska*

ARNOLD POWELL
*Columbus College*

Two experiments were carried out to determine the effect of blocked presentation on storage and retrieval. It was argued that blocking results in the simultaneous accessibility of information in working memory. Simultaneous accessibility of semantic relationships provides a context for encoding which contributes to the detection of interitem relationships and the storage of representations with regard to these relationships. In Exp. I, blocking the presentation of ambiguous words partially determined the meaning of words to be remembered. In Exp. II, it was found that recall differences were not predictable from detection differences alone. The results were interpreted as supporting what was called the simultaneous encoding hypothesis and a view of information processing which incorporates changes in discriminability, the integration of subjective units, and the development of retrieval plans.

Several studies of single-trial free recall of categorized word lists have found that recall was facilitated when the order of presentation was organized according to category structure (Cofer, Bruce, & Reicher, 1966; D'Agostino, 1969; Dallett, 1964; Puff,[3] 1966). Recall performance was generally superior in presentation conditions having the maximum number of category repetitions (C-Reps), i.e., blocked presentation, when compared with presentation conditions having minimum C-Reps, i.e., "random" presentation. Explanations of the blocking phenomena are closely related to basic assumptions concerning organization in free recall which have emphasized: (*a*) associative changes in permanent memory (Cofer et al., 1966; Puff, 1966), (*b*) the development of retrieval strategies based on the general structure of the list (Slamecka, 1968), and (*c*) more inclusive views which postulate an increase in discriminability of permanent memory representation, an integration of subjective units, and the development of retrieval plans (Bower, Clark, Lesgold, & Winzenz, 1969).

D'Agostino (1969) and Glanzer (1969) have proposed that during presentation, *S*s retrieve information from permanent memory which is used to relate to-be-remembered, TBR, words. The simultaneous accessibility of relational information in working memory provides a necessary condition for organizing TBR words in terms of these relationships, that is, *S*s establish a working memory context for encoding. When related words are presented in close proximity, i.e., not separated by unrelated words, information processing is facilitated. This view will be called the simultaneous encoding hypothesis.

The present research was directed at two issues: (*a*) Does blocked presentation provide a context for encoding? (*b*) Which view of information processing provides as adequate account of how *S*s use relationships that are made simultaneously accessible when the presentation is blocked?

## EXPERIMENT I[4]

Ambiguous words are characterized by a single phonological representation and multiple meanings. According to the simultaneous encoding hypothesis, the way in which an ambiguous word (or any word) is encoded will depend in part on the context of encoding. Bobrow and Light (reported by Bower, 1970) used adjectives in the input stages of free and cued recall experiments in order to direct the encoding process to semantic representations of the TBR words that could be used to organize the list of 44 words into eight categories. In the second group, the adjectives directed encoding to semantic representations that would enter into less optimal retrieval strategies. Free and cued recall were superior in the first of these two conditions. In Exp. I, it was hypothesized that maximizing C-Reps with ambiguous words would provide the context for coding and recoding ambiguous words according to accessible relationships, i.e., would determine the meaning of the words to be remembered.

---

[1] This research was supported in part by the University of Nebraska Research Council Grant G04-4128-44R awarded to the first author and a National Science Foundation Predoctoral Traineeship awarded to the second author.

[2] Requests for reprints should be sent to Richard M. Weist, Department of Psychology, 1132 Oldfather Hall, University of Nebraska, Lincoln, Lincoln, Nebraska 68508.

[3] C. Richard Puff has recently reviewed the literature in this area.

[4] A shorter version of this experiment which emphasized the serial position findings was presented at the Rocky Mountain Psychological Association, Denver, May 1971.

### *Method*

*Subjects.*—The *S*s were 92 male and female introductory psychology students at the University of Nebraska-Lincoln. The *S*s were tested in groups of 5–14. Experimental groups were run in a random order and *S*s were assigned to groups as they signed up to participate.

*Stimulus materials.*—Two kinds of 30-word categorized lists having five categories were obtained primarily from the Battig and Montague (1969) norms. The stimulus words were either ambiguous (Ambig) or nonambiguous (Non-A). Four of the five categories of the ambiguous words were selected from the norms and the fifth was intuitively generated. The words selected were categorized in this experiment with respect to their less salient meaning, e.g., "grasshopper" and "screwdriver," occurred in the norms as responses to "an insect" and "a carpenter's tool," and also as responses to "an alcoholic beverage." These words were treated as members of the category "an alcoholic beverage" in this study. The categorized word lists were then arranged into two presentation forms; the number of C-Reps was either maximized having 25 C-Reps or minimized having 0 C-Reps (Puff, 1966). Three randomizations of the four lists, i.e., Ambig 25 C-Reps, Non-A 25 C-Reps, Ambig 0 C-Reps, and Non-A 0 C-Reps, were prepared for presentation.

*Procedure.*—The *S*s received three free recall trials. Stimuli were presented at a 2.5-sec. rate on a Carousel slide projector equipped with a separate timing device. The *S*s were allowed 90 sec. to recall as many words as they could in any order they wished. The words recalled on each trial were written on the individual pages of test booklets.

*Design.*—C-Reps (0 vs. 25) and ambiguity (Ambig vs. Non-A) were between-*S*s' effects, with 23 *S*s in each condition. A serial position analysis was done on the number of words recalled on Trials 1–3, and 30 serial positions were collapsed into six blocks of five successive positions each. The results were evaluated in a $2 \times 2 \times 3 \times 6 \times 23$ *S*s analysis of variance. Clustering measures (ARC) were obtained (Roenker, Thompson, & Brown, 1971) and evaluated for the four lists over three trials in a $2 \times 2 \times 3 \times 23$ *S*s analysis of variance.

### *Results*

The organizational performance is shown in Fig. 1. When the Ambig's were presented with 0 C-Reps, there was no evidence of organization by *E*-defined categories in Trial 1; i.e., observed clustering did not exceed the expected value. The level of clustering was generally higher for Non-A's and maximizing C-Reps facilitated clustering with both Ambig's and Non-A's to a similar extent. The analysis of variance indicated significant main effects of ambiguity, $F\ (1,\ 88) < 94.60$, $p = .001$; C-Reps, $F\ (1,\ 88) = 185.86$, $p < .001$; and

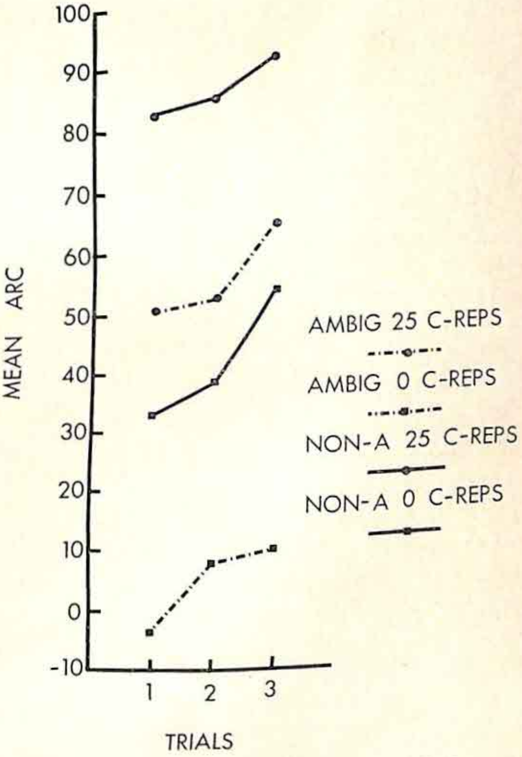


FIG. 1. Relative clustering (mean ARC) over trials for the four experimental conditions (Ambig 25 C-Reps, Ambig 0 C-Reps, Non-A 25 C-Reps, and Non-A 0 C-Reps).

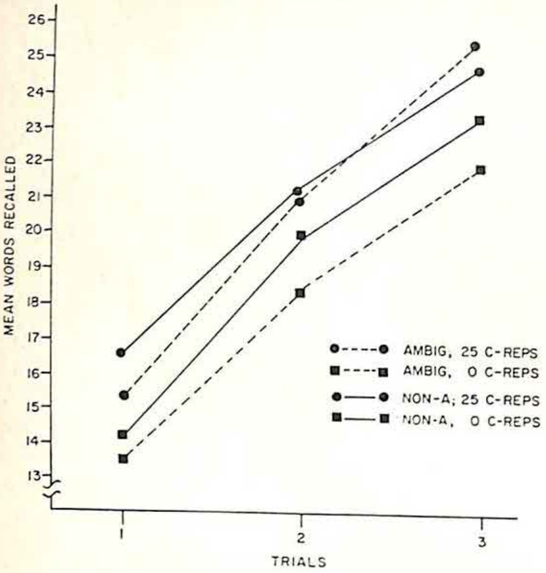


FIG. 2. Recall as a function of trials for ambiguous and unambiguous word lists presented with either 25 or 0 C-Reps.

trials, $F$ (2, 176) = 15.48, $p < .001$. None of the interactions were significant. Increased clustering with blocked presentations has generally been obtained. The most interesting finding was in the Ambig condition, where the blocked presentation directed encoding to organization that was not apparent in the 0 C-Rep condition.

Figure 2 shows that blocked presentation consistently facilitated recall over three trials. The main effect of C-Reps, $F$ (1, 88) = 10.53, $p < .01$; and trials, $F$ (2, 176) = 278.82, $p < .01$, were significant. The main effect of ambiguity and the C-Reps × Trials interaction were not significant. Serial position curves had a greater primacy than recency effect as has previously been found with organized lists (Deese & Kaufman, 1957). Position curves tended to flatten more slowly in the Ambig 0 C-Reps condition. This result was reflected in the C-Reps × Position × Trials interaction, $F$ (10, 880) = 3.57, $p < .001$; and Ambiguity × C-Reps × Position × Trials interaction, $F$ (10, 880) = 1.93, $p < .05$. In general, the results of Exp. I support the simultaneous encoding hypothesis and clearly show that blocked presentation has an effect on the storage of information in permanent memory.

## EXPERIMENT II

Puff's (1966) explanation of the blocking effect was based on the assumption of implicit representational and associative response priming. This position may be called the "strong storage" hypothesis because it makes no separate assumptions concerning retrieval plans. Previous research (Weist, 1972) found this position to be inadequate. Experiment II contrasts Slamecka's (1968) explanation of information processing under conditions of blocked and "random" presentations with the more inclusive explanations of Bower et al. (1969). Slamecka proposed that *S*s formulate a plan for retrieval during presentation and store TBR words independently. Maximizing input C-Reps on this view merely increases the likelihood of detecting the general category structure of the lists. Cofer et al. (1966) called this kind of position the "strong coding" hypothesis. In a multitrial free recall experiment, the strong coding hypothesis would predict that recall performance in maximum and minimum C-Rep presentation conditions should converge as *S*s detect the category structure of the list; and where there are no differences in *S*s' capacity to report the general list structure there should be no recall differences.

Two methodological preliminaries were dealt with in pilot studies. The first study was a multitrial replication of Puff (1966). A list having three categories (NC3) with 10 items per category (IPC10) and Puff's procedures for determining levels of C-Reps were used to generate Random, 0, 9, 18, and 27 C-Reps presentation conditions. Puff's finding that recall and clustering increase as a function of C-Reps was supported on Trial 1 but that data was too variable to detect these differences over 10 trials. On the basis of these results, it was decided that only maximum and minimum C-Reps conditions would be used in further investigation and both low-frequency and NC5 IPC6 lists would be added to the design.

In a previous study (Weist, 1972), 99% of the *S*s were able to report four out of four categories on NC4 IPC6 lists after five free recall trials. In the second pilot study, the progress of detection was plotted over five trials. Two NC5 IPC6 lists having either high- or low-frequency (HF or LF) category members (see Exp. II) were used. The C-Reps in presentation were either maximized (25 C-Reps) or minimized (0 C-Reps). Independent groups of *S*s which ranged in size from 17 to 22 received one of the following sequences of input (I) and output (O) free recall phases: I; IOIOI; IOIOIOIOI (i.e., 1I, 3I, or 5I). After the last input phase, *S*s were asked to name the five categories. In all conditions, 100% of the *S*s reported 4 of 5 categories correctly after 3I phases. In the HF conditions, *S*s in the 25 C-Rep condition held an early detection advantage over 0 C-Rep conditions: 83% versus 62% on 1I, 100% versus 80% on 3I, and 100% versus 95% on 5I. A few 5I *S*s used "animal" inclusively to incorporate "insects" and "four-legged animals." In the LF groups, there were no differences in the detection of list structure in the 25 and 0 C-Rep conditions: 53% versus 59% on 1I, 84% versus 82% on 3I, and 96% versus 94% on 5I.

If, as the strong coding hypothesis suggests, determining the general list structure is the only factor contributing to the facilitation of recall found in single-trial studies when C-Reps are maximized, the following results are expected: (*a*) Given an NC3 IPC10 list, there should be a small Trial 1 difference which is no longer obtained by Trial 3, (*b*) given an NC5 IPC6 LF list, there should be no difference between 0 and 25 C-Rep conditions, (*c*) given an NC5 IPC6-HF list, there should be an early-trial difference between 0 and 25 C-Reps and a convergence by Trial 5. These predictions were tested in Exp. II.

### *Method*

*Subjects*.—The *S*s were 112 introductory psychology students at the University of Nebraska-Lincoln and the same population used in Exp. I and the pilot studies. The *S*s were run in groups of 5–11. Experimental groups were tested in a random order and *S*s were assigned to groups in the order in which they volunteered to participate.

*Stimulus material*.—Four 30-word lists were used. The lists differed in list structure and frequency, having either 3 or 5 categories (NC3 or NC5) and having relatively high- or low-frequency (HF or LF) responses to category labels. According to the Battig and Montague (1969) norms, the average cultural frequencies for the high-frequency words were 111.8, 93.5, and 160.7 for each of the categories in the NC3 list, and 68.8, 72.1, 63.8, 98.7, and 132.2 for the NC5 list. The average cultural frequencies for the low-frequency words were 10.3, 9.8, and 8.0 for the NC3 list and 10.8, 11.0, 4.0, 11.7, 12.8 for the NC5 list. Five presentation orders were generated for each of the four lists at the maximum (Max) and minimum (Min) level of C-Reps. The Max C-Reps was 27 on the NC3 lists and 25 on the NC5 lists. The Min C-Reps was 0 for both levels of the list structures. These were the same lists used in the pilot studies.

*Procedure*.—The procedure was the same as Exp. I except that five free recall trials were used.

*Design*.—A serial position analysis was done on the number of words recalled. The 30 serial positions were collapsed into six blocks of five successive positions each and the results evaluated in a 2 × 2 × 2 × 5 × 6 × 14 *S*s analysis of variance. C-Reps (Max vs. Min), list structure (NC3 vs. NC5), and word frequency (HF vs. LF) were between-*S*s treatments, with 14 *S*s in each group, and five trials and six positions were within-*S*s treatments.

### *Results*

Figure 3 shows that there was a consistent facilitation due to maximizing C-Reps over five trials in all four groups, and there was no indication of converging functions. The analysis of variance indicated that all of the main effects except list structure were significant: frequency, $F$ (1, 104) = 9.86, $p < .01$; C-Reps, $F$ (1, 104) = 43.41, $p < .001$; trials, $F$ (4, 416) = 384.00, $p < .001$; and position, $F$ (5, 520) = 6.93, $p < .001$. Each of the four groups were analyzed independently so that the results could not be attributed to inflated degrees of freedom and because differential predictions were made. The effect of maximizing C-Reps is always significant, HF-NC3; $F$ (1, 26) = 4.33, $p < .05$; LF-NC3, $F$ (1, 26) = 13.43, $p < .01$; HF-NC5, $F$ (1, 26) = 7.24, $p < .025$; and LF-NC5, $F$ (1, 26) = 23.35, $p < .001$. The C-Reps × Trials interaction was significant only in the LF-NC5 group where there was a slight divergence, $F$ (4, 104) = 3.02, $p < .05$. On the basis of the strong coding hypothesis, Max and Min recall functions should converge more rapidly when there

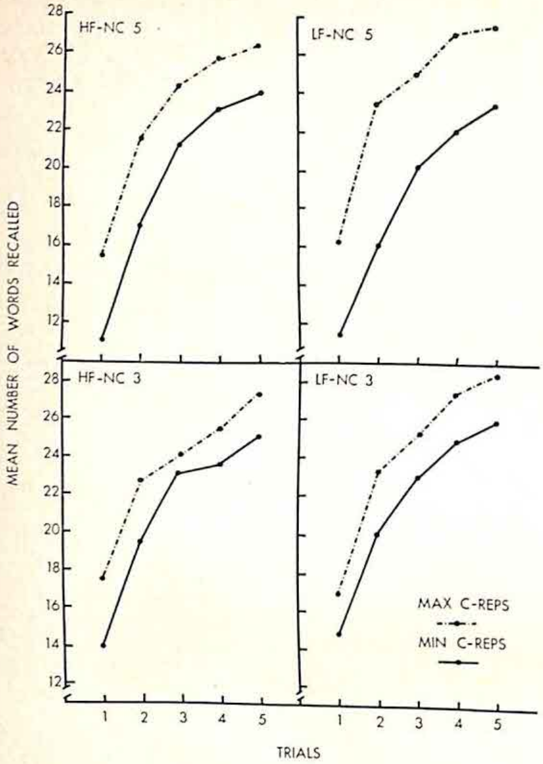


FIG. 3. Recall as a function of trials for the Max and Min C-Reps list conditions for the four experimental groups combining two levels of word frequency HF and LF with two levels of list structure NC3 and NC5.

are fewer categories to detect (NC3 vs. NC5). The data show no indication of converging recall trends in any of the conditions. While there were no differences in *S*s' capacity to detect the list structure of the LF-NC5IPC6 list in the second pilot study, the largest recall differences were found in this condition in Exp. II. When detection differences were found to converge over five trials (HF-NC5IPC6), there was no corresponding convergence in the recall functions. To the extent to which the capacity to report the categorical structure of the list reflects the development of retrieval plans, these findings show that the development of retrieval plans alone does not adequately explain the blocking effect.

The Position × Trials interaction, $F$ $(20, 2080) = 5.88$, $p < .001$, was significant indicating that the position curves flatten with trials. The position curves had a greater primacy than recency effect supporting the results in Exp. 1. There was no C-Reps × Position interaction, which supports Glanzer's (1969) argument that blocking affects both LTS and STS.

## DISCUSSION

In cued encoding and retrieval studies (Thomson & Tulving, 1970; Tulving & Olser, 1968), it was demonstrated that the way in which information is remembered depends in part on the encoding context. In free recall studies, representations in working memory provide the context for encoding. D'Agostino (1969) proposed that given a random presentation, *S*s retrieve prior information which is needed to determine organization. In a study which directly supports this hypothesis, Rundus (1970) found that *S*s tended to rehearse words from the same category together rather than shift categories. Glanzer (1969) found that decreasing the distance between related words increases *S*s' capacity to make critical relationships simultaneously accessible in working memory. Blocking presentation leads to the most efficient use of encoding context (Puff, 1966) and retrieval cues are more effective under blocked than random conditions (Lewis, 1971).

The results of Exp. I and II are supported by previous research which has also found the strong coding hypothesis to be inadequate. Informing *S*s about the general structure of a categorized list (D'Agostino, 1969) did not decrease the effectiveness of the blocked presentation as the strong coding hypothesis would predict. The use of recognition tasks to eliminate the retrieval stage of free recall reduced but did not eliminate the effect of blocking (Bower et al., 1969; D'Agostino, 1969). D'Agostino and Lewis (1971) found that blocking increases the number of items recalled from a category rather than the number of categories recalled. These findings indicate that blocking facilitates the discriminability of permanent memory representations and the integration of subjective units.

The serial position data from both Exp. I and II support some of the results and conclusions of Deese and Kaufman (1957) concerning the effect of list organization upon serial position curves. Deese and Kaufman found that there was some flattening of the serial position curve and a shift from a greater recency effect to a greater primacy effect as the sequential organization of the words to be recalled more closely approximated the sequential organization of English sentences. These effects were attributed to the sequential organization, per se, but the results of the present study suggest that they may be due, more generally, to the presence of any organization. When blocked presentation has caused a C-Reps × Position interaction, the effect was limited to the middle portion of the list (Cole, Frankel, & Sharp, 1971; and the Ambig condition of Exp. I). These findings also support Deese and Kaufman and indicate a faster rate of flattening in the more organized lists.

In general, the present research indicates that *S*s interrelate category members at least during the presentation phase of a free recall experiment. The presentation of related words in close proximity provides a context for encoding TBR words in terms of accessible relationships. As Bower et al. (1969) suggest, an adequate account of the information processing that occurs in free recall must include an analysis of changes in discriminability of TBR information (e.g., Kintsch, 1970), an account of the integration of subjective units (e.g., Tulving, 1964), as well as an understanding of the development of retrieval plans (Slamecka, 1969).

## REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories. A replication and extension of the Connecticut norms. *Journal of Experimental Psychology*, 1969, **80**, 1–46

BOWER, G. H. Organizational factors in memory. *Cognitive psychology*, 1970, **1**, 18–46.

BOWER, G. H., CLARK, M. C., LESGOLD, A. M., & WINZENZ, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 323–343.

COFER, C. N., BRUCE, D. R., & REICHER, G. M. Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 1966, **71**, 858–866.

COLE, M., FRANKEL, F., & SHARP, D. Development of free recall learning in children. *Development Psychology*, 1971, **4**, 109–123.

D'AGOSTINO, P. R. The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 815–820.

DALLET, K. M. Number of categories and category information in free recall. *Journal of Experimental Psychology*, 1964, **68**, 1–12.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, 1957, **54**, 180–187.

GLANZER, M. Distance between related words in free recall: Trace of the STS. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 105–111.

KINTSCH, W. Models of free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

LEWIS, M. Q. Categorized lists and cued recall. *Journal of Experimental Psychology*, 1971, **87**, 129–131.

PUFF, C. R. Clustering as a function of the sequential organization of word lists. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 503–506.

ROENKER, D. L., THOMPSON, C. A., & BROWN, S. C. A comparison of measures for estimation of clustering in free recall. *Psychological Bulletin*, 1971, **76**, 45–48.

RUNDUS, D. J. An analysis of rehearsal processes in free recall. (Institute for Mathematical Studies in the Social Studies, Tech. Rep. No. 159) Palo Alto, Calif: Stanford University, 1970.

SLAMECKA, N. J. An examination of the trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

THOMSON, D. M., & TULVING, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, **86**, 255–262.

TULVING, E. Intratrial and intertrial retention: Notes toward a theory of free recall verbal learning. *Psychological Review*, 1964, **71**, 219–237.

TULVING, E., & OSLER, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593–601.

WEIST, R. M. Associative structure in free recall. *Journal of Experimental Psychology*, 1972, in press.

(Received October 22, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 404–409

# PROBABILITY DISCRIMINATION INDICATED BY STIMULUS PREDICTIONS AND REACTION SPEED: EFFECTS OF S–R COMPATIBILITY[1]

E. SCOTT GELLER,[2] CHARLES P. WHITMAN, AND JOHN C. FARRIS

*Virginia Polytechnic Institute and State University*

Independent effects of stimulus probability and prediction outcome on discrete two-choice reaction speed (CRS) were determined for a compatible and an incompatible condition. A symbol occurred at a "left" location on 70% of the initial 150 trials; then the relative frequencies were reversed so that the symbol occurred on the "right" for 70% of 150 presentations. Each trial consisted of *S*s' verbal prediction, a stimulus presentation, and *S*s' choice response. Group C (compatible) pulled a left-hand trigger to identify "left" and a right-hand trigger to identify "right." The reverse S–R mapping arrangement was used for Group I (incompatible). The degree of probability discrimination reflected by *S*s' stimulus predictions and reaction speeds was greater for Group C than Group I. The marked effects of prediction outcome on CRS did not interact with compatibility. The relative importance of perceptual and response bias in determining effects of CRS are discussed.

Investigators have varied the compatibility of the stimulus–response (S–R) relationship in a choice reaction time (RT) paradigm by manipulating the set of stimulus alternatives paired with a particular response set or by changing the "naturalness" of the pairings between the stimulus and response alternatives (see review by Smith, 1968). Such research has demonstrated the influence of the compatibility variable on the relationship between choice RT and several other independent variables. For example, the slope of the linear relationship between choice RT and $\log_2$ of the number of stimulus alternatives was an inverse function of the degree of S–R compatibility (Smith, 1968; Welford, 1960); the "repetition effect" (i.e., faster RTs to stimulus repetitions than alternations) was greater for more complex S–R arrangements (Bertelson, 1963; Schvaneveldt & Chase, 1969); and the effect of changes in stimulus probability on choice RT varied inversely with the degree of S–R compatibility (Fitts, Peterson, & Wolpe, 1963; Sanders, 1970). In addition, two recent choice RT studies which required *S*s to predict the stimulus presentations found the "prediction outcome effect" (i.e., faster RTs to correct anticipations) to be more prominent in an incompatible than in a compatible situation (Keele, 1969; Whitman & Geller, 1971). The present choice RT experiment studied the influence of S–R compatibility on both probability and prediction outcome effects when the relative frequencies of the two stimulus alternatives were reversed midway through the experimental session. The manipulation of compatibility was identical to that of several other choice RT experiments (e.g., Wallace, 1971; Whitman & Geller, 1971) and was accomplished by varying the mapping between left and right stimulus lights and left-hand and right-hand response triggers.

Theoretical analyses of the choice RT process assume that at least two cognitive processes are involved: stimulus identification and response selection. Using the two-stimulus, three-response design introduced by Laberge and Tweedy (1964), researchers have studied the relative importance of the perceptual and response components when choice RT is affected by changes in stimulus probability or prediction outcome. A majority of studies have implicated the stimulus identification

---

[1] This research was partially supported by the Small Grants Program of the College of Arts and Sciences, Virginia Polytechnic Institute and State University. Special assistance in collecting data was provided by David S. Post.

[2] Send reprint requests to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

process to a greater extent than the response selection process as the source of variance due to changes in probability or prediction outcome. That is, changes in stimulus probability rather than response probability determined most of the "probability effect" (Bertelson & Tisseyre, 1966; Hawkins & Hosking, 1969; Hawkins, Thomas, & Drury, 1970), and correct anticipation of the stimulus rather than the response resulted in the prediction outcome effect (Hinrichs & Krainz, 1970; Whitman & Geller, 1972).

The complexity of the S–R relationship in a choice RT situation has been shown to influence the importance of the response selection component of the choice reaction process. More specifically, in a highly compatible task reaction latencies are said to be largely determined by the perceptual component of the choice RT process, whereas response effects become more important determinants of choice RT when the S–R relationship is made less compatible (e.g., Sanders, 1970; Welford, 1960). Given this assumption, the observations that less compatible S–R arrangements resulted in a greater influence of stimulus probability on choice RT (e.g., Fitts et al., 1963; Sanders, 1970) and a larger effect of correct or incorrect stimulus anticipations on choice RT (Keele, 1969; Whitman & Geller, 1971) implicate response factors in the probability and prediction outcome effects. Thus, the research methodology used to study the relative roles of perceptual and response bias in a discrete choice RT task (i.e., three-stimulus, two-response design vs. compatibility manipulations) has determined whether or not the response selection process was significantly influenced by changes in probability or prediction outcome. The present experiment compared the effects of both probability and prediction outcome between high and low S–R compatibility conditions. A reliably larger influence of either variable on choice RT in the low compatible condition would necessarily implicate appropriate response effects.

## METHOD

*Subjects.*—Two experimental groups of 10 males and 10 females each participated in individual 1-hr. sessions. All *Ss* were enrolled in the introductory psychology course at Virginia Polytechnic Institute and State University; none had previously been in a reaction time experiment.

*Stimuli.*—For each of 300 stimulus presentations, the symbol ∩ was presented by the illumination of one of two digital readouts, each 5 × 1 cm. and separated on a horizontal axis by 35 cm. For each of the first three consecutive blocks of 50 trials, 70% of the presentations (i.e., 35) were on the "left" and 30% (15) were on the "right." Then the relative frequencies of the left and right locations were reversed so that for each of the next three consecutive 50-trial blocks 70% of the stimuli were "right." The number of consecutive identical locations never exceeded seven for the more probable location or three for the less probable location.

*Procedure.*—Prior to each stimulus presentation, *Ss* verbally predicted whether they expected the left or right alternative. After approximately 2 sec., a ready buzzer sounded; and following a uniformly random time interval ranging from .5 to 2 sec., the left or right readout was illuminated. The *Ss* identified the stimulus location (left or right) as quickly as possible by depressing one of two microswitches which were positioned within spatially separated reaction triggers, the left trigger being gripped by *Ss*' left hand and the right trigger by *Ss*' right hand.

The *Ss* of the compatible condition (Group C) pressed the left trigger following a left stimulus and the right trigger following a right stimulus. The reverse S–R mapping relationship was in effect for *Ss* in the incompatible condition (Group I); i.e., the right trigger was pressed to identify left and the left trigger was pressed to identify right. A choice reaction turned off the stimulus and stopped a Hunter Klockounter, Model D. The latency between each stimulus presentation and an identification response was measured to the nearest millisecond. Anticipatory responses and inappropriate choices were reported to *E* via indicator lights.

The task instructions requested that *Ss* concentrate on each stimulus prediction and emphasized the importance of both speed and accuracy when identifying stimulus presentations. The *Ss* were informed that the stimulus presentations were programmed on punched paper tape, but were not told anything about the intrasequential distribution of the two stimulus locations. At the start of each individual session, 10 practice trials were given during which *Ss*' questions concerning their task were answered.

## RESULTS

*Compatibility and reaction speed.*—Each latency was converted to a choice reaction speed (CRS = 1/sec). The errors (i.e., anticipatory responses or inappropriate identifications) were eliminated from the data analysis and did not exceed 4% (i.e., 12) for any given *S*. The independent effects of prediction outcome and relative frequency on CRS were determined for each compatibility condition by categorizing the speeds of each *S* according to the stimulus presentation (left or right location) and the prediction outcome determined by that stimulus presentation (correct or incorrect prediction), calculating the category means for each *S* over 50-trial blocks, and determining group means for each category. For all *S*s, each category mean included at least four data entries. For the Compatible (Group C) and the Incompatible (Group I) conditions, Fig. 1 depicts the average CRS following correctly predicted "lefts," correctly predicted "rights," incorrectly predicted "lefts," and incorrectly predicted "rights" for consecutive blocks of 50 trials. A between-group speed comparison for each of the four categories indicated that the speeds of Group C were markedly faster than those of Group I. For both groups, average CRS was faster to correctly predicted events, and for the prereversal trial blocks CRS was faster to the more probable stimulus alternative of each prediction outcome category. Following the probability reversal, only Group C demonstrated a complete speed reversal within both the correctly predicted and incorrectly predicted event categories.

The overall analysis of variance, a 2 (Groups) × 2 (Prediction Outcomes) × 2 (Stimuli) × 2 (Prereversal vs. Postreversal) ×3 (Trial Blocks) factorial, demonstrated significant main effects of group, $F\,(1, 38) = 9.62$, $p < .005$; prediction outcome, $F\,(1, 38) = 78.97$, $p < .001$; stimulus, $F\,(1, 38) = 43.62$, $p < .001$; reversal $F$, $(1, 38) = 16.14$, $p < .001$; and trial block, $F\,(2, 76) = 12.50$, $p < .001$. Of the interactions only four reached significance, i.e., Stimulus × Reversal, $F\,(1, 38) = 30.42$, $p < .001$; Group × Stimulus × Reversal, $F\,(1, 38) = 5.48$, $p < .025$; Stimulus × Trial Block, $F\,(2, 76) = 4.04$, $p < .025$; and Group × Stimulus × Reversal × Trial Block, $F\,(2, 76) = 4.03$, $p < .025$. These significant interactions indicate that CRS was significantly influenced by the probability reversal and that the effect of the reversal on CRS was dependent on compatibility. To further examine the compatibility effects a 2 (groups) × 2 (prediction outcomes) × 2 (stimuli) × 3 (trial

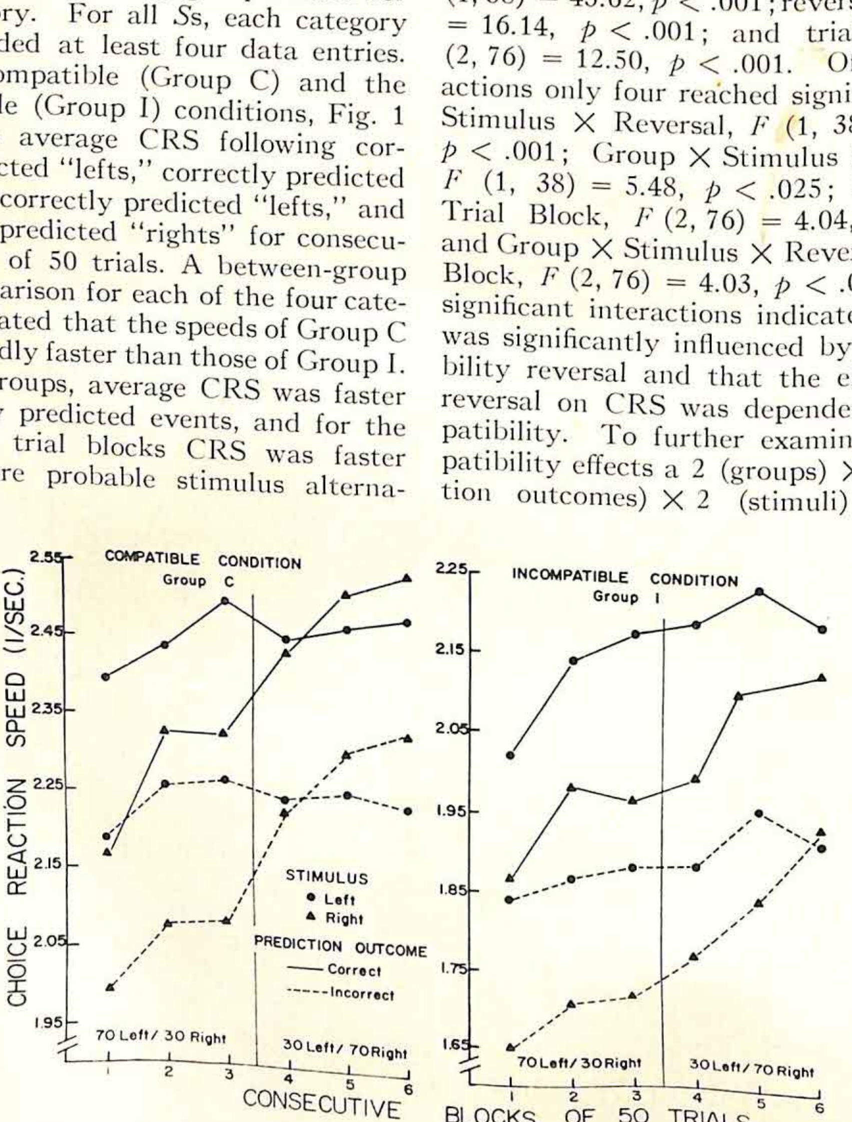


FIG. 1. Choice reaction speed as a function of stimulus probability and prediction outcome for compatible (Group C) and incompatible (Group I) stimulus–response mappings.

blocks) factorial was calculated for both prereversal and postreversal.

The prereversal analysis indicated significant main effects of group, $F$ (1, 38) = 7.36, $p < .01$; prediction outcome, $F$ (1, 38) = 80.55, $p < .001$; stimulus, $F$ (1, 38) = 63.34, $p < .001$; and trial block, $F$ (2, 76) = 6.97, $p < .005$. None of the interactions were significant (all $p$s > .25). For postreversal data, significant main effects of group, $F$ (1, 38) = 11.54, $p < .005$; prediction outcome, $F$ (1, 38) = 61.29, $p < .001$; and trial block, $F$ (2, 76) = 7.00, $p < .005$, were obtained. The main effect of stimulus was not significant, $F$ (1, 38) = 2.45, $.25 > p > .10$. Of the interactions, significance was reached for Group × Stimulus, $F$ (1, 38) = 10.95, $p < .005$; and Stimulus × Trial Block, $F$ (2, 76) = 6.22, $p < .005$.

*Compatibility and probability learning.*—To study the effect of S–R compatibility on *S*'s prediction strategies, the proportion of *S*s in each group who predicted a left stimulus was calculated for successive 50-trial blocks. As indicated in Fig. 2, for each trial block Group C predicted the more frequently occurring stimulus alternative more often than Group I. That is, for Group C the proportion of left predictions was higher for the first three trial blocks and was lower following probability reversal. The analysis of variance for the proportion of left predictions as a function of compatibility, reversal, and trial block (i.e., a 2 × 2 × 3 factorial) indicated a significant main effect of only reversal, $F$ (1, 38) = 23.83, $p < .001$. Significant interactions obtained were Group × Reversal, $F$ (1, 38) = 6.29, $p < .005$, and Reversal × Trial Block, $F$ (2, 76) = 20.26, $p < .001$. Post hoc $t$ tests between the trial block proportions of each group ($df = 19$) indicated that the differences on Trial Blocks 1, 5, and 6 were significant beyond the .05 level (two tailed).

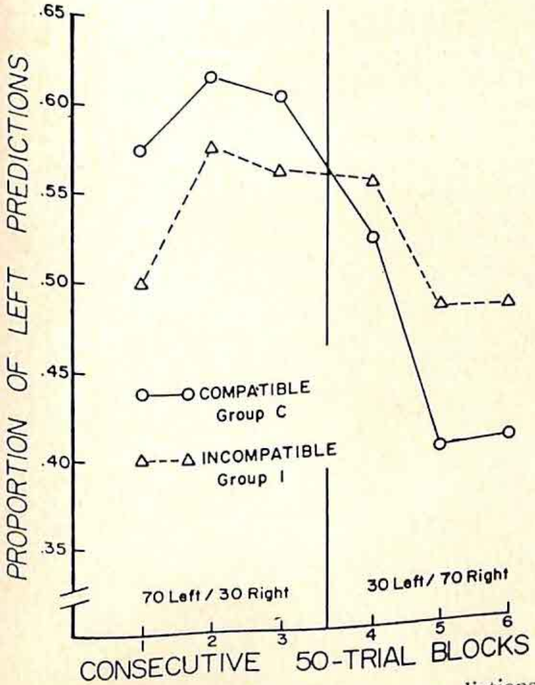


FIG. 2. Proportion of "left" stimulus predictions for Groups C and I as a function of consecutive blocks of 50 trials.

## DISCUSSION

The results indicated that both dependent variables, choice reaction speed (CRS) and the proportion of "left" predictions, were significantly influenced by the degree of S–R compatibility. As expected, choice reactions were much faster when the spatial attribute of each stimulus alternative (left or right) was identical to that of its appropriate identification response (cf. Wallace, 1971). When the relative frequencies of the left and right stimulus locations were reversed, CRS to both correctly and incorrectly predicted events reflected the reversal to a greater extent when the S–R arrangement was highly compatible, (i.e., Group C). Theoretically, the additional response bias introduced by the incompatible S–R mapping (i.e., Group I) inhibited the postreversal probability discrimination as indicated by differential speeds to each stimulus alternative. Moreover, weaker probability discrimination by *S*s of Group I was demonstrated by the analysis of *S*'s prediction strategies. That is, for both prereversal and postreversal, *S*s of Group C predicted the more frequently occurring stimulus alternative reliably more often than did *S*s of Group I. Apparently, requiring *S*s to identify a stimulus location with a response in the contralateral location added task complexity sufficient to inhibit the probability discrimination implied by *S*s' predictions as well as their reaction speeds.

It is noteworthy that the speed reversals reported in the study which introduced probability reversal in a choice RT paradigm (Geller, Whitman, Wrenn, & Shipley, 1971) were less complete than those of Group C in the present study and more complete than those of Group I. In the experiment by Geller et al., a discrimination of the relative frequencies of an incorrectly predicted event was implied only to the extent that CRS was significantly influenced by stimulus probability during prereversal but not during postreversal. For the incorrectly predicted events of the present study a marked speed reversal was observed for Group C, but the speeds of Group I were not influenced by the probability reversal until the last 50-trial block of postreversal. Since Geller et al. varied the stimulus alternatives on a shape dimension (i.e., *U* or ∩ symbols in the center of a screen) and the response alternatives on a location dimension (i.e., left or right response triggers), the degree of compatibility in that study was less than that of Group C in the present study but greater than that of Group I in the present study. Thus, comparisons between the probability reversal effects in these studies support the notion of a continuous compatibility dimension (cf. Whitman & Geller, 1972). Specifically, combining the present results with those of Geller et al., the degree of probability discrimination reflected by differential speeds to two stimulus alternatives was a direct function of the level of S–R compatibility.

The observation that the effect of correct or incorrect stimulus predictions on choice RT was not influenced by a decrease in S–R compatibility indicated that the prediction outcome effect is primarily a perceptual phenomenon. A similar conclusion resulted from previous choice RT research using a three-stimulus, two-response paradigm (Hinrichs & Krainz, 1970; Whitman & Geller, 1972). Likewise, the present observations that the probability effect was not influenced by the compatibility condition during prereversal and was less prominent for the incompatible condition during postreversal supported the conclusion of the three-stimulus, two-response experiments: that changes in relative stimulus frequency influence primarily the identification component of the choice RT process (Bertelson & Tisseyre, 1966; Hawkins & Hosking, 1969; Hawkins et al., 1970). However, these findings of the present study were inconsistent with the results of earlier experiments which varied compatibility and stimulus probability in a choice RT paradigm (Fitts et al., 1963; Sanders, 1970). Indeed, these investigators found probability changes to have an increasingly greater effect on choice RT as the degree of S–R compatibility decreased and thus showed response factors to be important determinants of the probability effect. Although this discrepancy may be attributed to any one of a number of procedural variations between these studies (e.g., operational definitions of compatibility, number of stimulus alternatives, the occurrence of a probability reversal), it would be most provocative to consider a specific difference between the task instructions of the present study and those of the other experiments. Specifically, Fitts et al. and Sanders informed their *S*s of the relative frequency of occurrence for each stimulus alternative, but *S*s in the present experiment were not informed of the stimulus probabilities. Consequently, the dependent variables of the present study (i.e., choice latency and stimulus predictions) were indications of *S*s' learning to discriminate between stimulus probabilities. However, when *S*s knew the relative frequency of each stimulus at the start of the experiment, it is reasonable to assume that choice latencies did not reflect discrimination learning to that same extent, and consequently different effects of S–R compatibility or task complexity might be expected.

The importance of the amount of probability information given *S*s in a choice RT experiment is also evident when one considers that researchers who advised *S*s of the stimulus probabilities prior to the experimental session observed an interaction between prediction outcome and compatibility (Keele, 1969) or between prediction outcome and probability (Geller & Pitz, 1970; Hinrichs & Craft, 1971), but that such interactions were not observed when *S*s' instructions did *not* include information about the stimulus probabilities (i.e., neither interaction was obtained in the present study).

Thus, a worthwhile goal for further research would be to derive an explanation for the apparently discrepant roles of response bias in the two discrete RT paradigms (i.e., three stimulus–two response versus relatively incompatible one-to-one mapping of stimulus and response). The results of the present study suggest that inconsistencies might be due to variations in the amount of probability information given *S*s during the task instructions.

## REFERENCES

Bertelson, P. S–R relationships and reaction time to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, 1963, **65**, 478–484.

Bertelson, P., & Tisseyre, F. Choice reaction as a function of stimulus versus response relative frequency of occurrence. *Nature*, 1966, **212**, 1069–1070.

Fitts, P. M., Peterson, J. R., & Wolpe, G. Cognitive aspects of information processing: II. Adjustments to stimulus redundancy. *Journal of Experimental Psychology*, 1963, **65**, 423–432.

Geller, E. S., & Pitz, G. F. Effects of prediction, probability, and run length on discrete choice reaction speed. *Journal of Experimental Psychology*, 1970, **84**, 361–367.

Geller, E. S., Whitman, C. P., Wrenn, R. F., & Shipley, W. G. Expectancy and discrete reaction time in a probability reversal design. *Journal of Experimental Psychology*, 1971, **90**, 113–119.

Hawkins, H. L., & Hosking, K. Stimulus probability as a determinant of discrete choice reaction time. *Journal of Experimental Psychology*, 1969, **82**, 435–440.

Hawkins, H. L., Thomas, G. B., & Drury, K. B. Perceptual versus response bias in discrete choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 514–517.

Hinrichs, J. V., & Craft, J. L. Verbal expectancy and probability in two-choice reaction time. *Journal of Experimental Psychology*, 1971, **88**, 367–371.

Hinrichs, J. V. & Krainz, P. L. Expectancy in choice response time: Anticipation of stimulus or response? *Journal of Experimental Psychology*, 1970, **85**, 330–334.

Keele, S. W. Repetition effect: A memory-dependent process. *Journal of Experimental Psychology*, 1969, **80**, 243–248.

LaBerge, D., & Tweedy, J. R. Presentation probability and choice time. *Journal of Experimental Psychology*, 1964, **68**, 477–481.

Sanders, A. F. Some variables affecting the relation between relative stimulus frequency and choice reaction time. *Acta Psychologica*, 1970, **33**, 45–55.

Schvaneveldt, R. W., & Chase, W. G. Sequential effect in choice reaction time. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

Wallace, R. J. S–R compatibility and the idea of a response code. *Journal of Experimental Psychology*, 1971, **88**, 354–360.

Welford, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–229.

Whitman, C. P., & Geller, E. S. Prediction outcome, S–R compatibility, and choice reaction time. *Journal of Experimental Psychology*, 1971, **91**, 299–304.

Whitman, C. P., & Geller, E. S. Prediction outcome and choice reaction time: Stimulus versus response anticipation. *Journal of Experimental Psychology*, 1972, in press.

(Received October 23, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 410–416

# MEMORY SEARCH AND THE RECALL OF DIGIT SERIES[1]

JAMES FRITZEN[2]

*Ohio State University*

Three experiments were conducted to examine how *S*s make use of repetitions in recalling digit strings. Each string was presented in chunks of two or three digits. In Exp. I, either the first, third, or fifth chunk was repeated across trials. Under conditions of immediate recall, only the first chunk improved. Under conditions of delayed recall, improvement was found in all three conditions. In Exp. II and III, repeated strings had chunks in common with one or two previous strings. The delayed recall data indicated that *S*s could make use of past repetitions from two separate previous strings. It was suggested that two separate accounts of repetition effects may be needed for immediate versus delayed recall.

Melton (1963) has demonstrated that the recall of a random string of digits will improve with repetition despite the intervening presentation and recall of as many as five other digit strings. One way to account for the effectiveness of past repetitions is in terms of a memory search process which terminates in the retrieval and utilization of past traces identical to the just-presented string.

Bower and Winzenz (1969) have proposed the reallocation hypothesis which entails a search through groups of strings in memory rather than individual strings. Strings that have certain digits (e.g., the first few digits) in common are grouped together or allocated to the same memory location. Search proceeds by checking addresses of locations, the address of each location being the first few digits of the string.

In agreement with the reallocation hypothesis, it is found that the repetition of only the middle or terminal portions of a digit string does not result in improved recall (Bower & Winzenz, 1969). Recall does improve, however, when partial repetition involves the first few digits of the string. Apparently, partial repetition is sometimes fruitless because the middle and terminal portions of strings, while they may be stored within memory locations, are not contained in the stored addresses of the strings.

Experiment I is essentially a replication of the Bower and Winzenz (1969) partial repetition effect, and is designed to explore the possibility that some factor other than reallocation might be responsible for the effect. In particular, during the presentation of a sequence of digits, there may be a tendency for *S*s to store the initial portion of the digit sequence in long-term store (LTS) and a tendency to store the latter portion of the sequence mainly (but not exclusively) in short-term store (STS). A relatively small degree of LTS strength for certain portions of the digit string under conditions of immediate recall might explain the lack of improvement for those portions when they are repeated.

In Exp. I, half of the *S*s recalled each string immediately after it was presented, the same procedure as that of Bower and Winzenz (1969). For those *S*s, it was predicted that the partial repetition effect of Bower and Winzenz would be replicated. The other half of the *S*s had a 15-sec. filler task between the presentation of every string and recall. Under those conditions, where *S*s cannot rely on STS for recall, it was hypothesized that the distribution of LTS strength across all portions of a digit string would be more nearly equal, resulting in improved recall when any portion of a digit string is repeated. The reallocation hypothesis, however, would predict the same effect for immediate and delayed recall; i.e., repetition of the initial portion of a string will improve recall while repetition of only the middle or terminal portions will have no effect.

---

[1] This paper is based on a dissertation submitted to the Ohio State University in partial fulfillment of the requirements for the PhD degree. The research reported here was supported by Grant MH 08526 from the National Institute of Mental Health, United States Public Health Service, and sponsored, in part, by the National Science Foundation through Grant GN 534.1 from the Office of Scientific Information Science Research Center, Ohio State University. The author wishes to express his gratitude to Neal F. Johnson for his suggestions and criticisms during the course of the research, and also to the members of the examination committee, J. R. Erickson, D. D. Wickens, and M. Jones.

[2] Requests for reprints should be sent to James D. Fritzen, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

## EXPERIMENT I

### *Method*

*Subjects.*—One hundred and twenty introductory psychology students served as *S*s. They were successively assigned to one of the six independent groups as they came to the experiment.

*Design.*—Each *S* had 20 recall tests on strings of 13 digits. Each string was segmented into five 2- or 3-digit chunks. For Groups I-1, I-3, and I-5, the recall test occurred immediately after the presentation of each string. The first, third, or fifth chunks, respectively, of every other string were the same 3 digits. For Groups D-1, D-3, and D-5, recall was delayed by a 15-sec. filler task immediately after the presentation of the string.

Across the six groups, the same constant chunks were used equally often. Within each group, the constant chunk occurred on the odd-numbered tests for one-half of the *S*s and on the even-numbered tests for the other half. The strings in which the constant chunks appeared were designated repeat or R-strings and the remaining strings were designated noise or N-strings.

*Materials.*—The strings were constructed from a table of random numbers using the digits 1–9 such that each digit could appear zero, once, or twice within a string, with the restriction that no digit could immediately repeat itself within the string. Each N-string was randomly assigned a different segmentation of three, three-digit groups and two, two-digit groups. Segmentation of the R-strings was constrained by the restriction that each string have a three-digit group in common; however, the same segmentation was never used more than twice. Five different lists were constructed as a precaution against list effects.

*Procedure.*—Each of the 20 strings was presented in five groups on five successive slides at a 1.5-sec. rate as *S* read the number on each slide, e.g., 321 was read as "three twenty-one." For the immediate recall condition, *E* said "recall" after the last slide of each string was shown. For the delay condition, recall occurred after a 15-sec. filler task which involved shadowing consonants at a .75-sec. rate (Bartz, 1969).

The *S*s were allowed 20 sec. to record their responses in an answer booklet and were encouraged to record their answers from beginning to end. Each page in the booklet had 13 evenly spaced dashes. Prior to the experimental trials, three practice trials were given involving recall of consonants.

### *Results*

The mean number of digits recalled from the constant chunks across trials is presented in Fig. 1 for each of the six groups. The improvement of constant first chunks was found to be significant, $F$ (9, 342) $= 6.16, p < .001$, and the improvement did not interact with delay, $F < 1$.

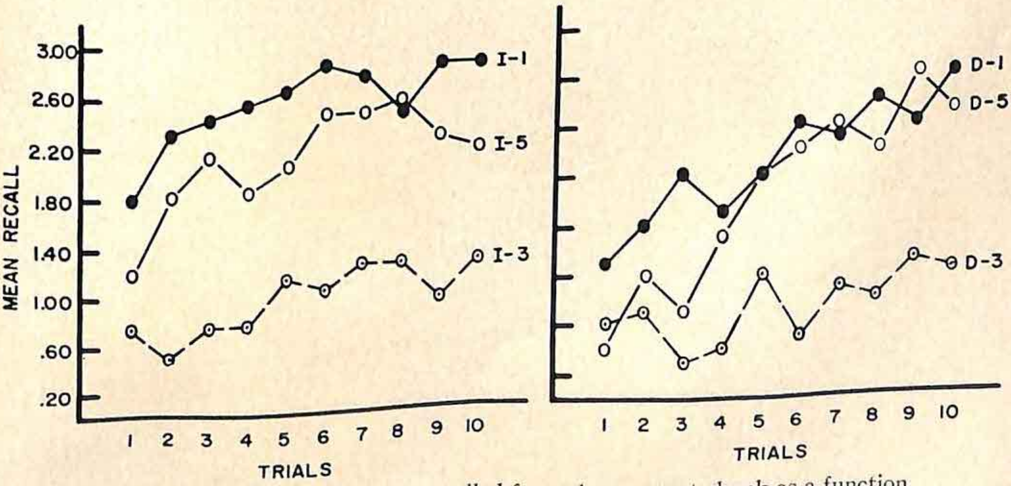


FIG. 1. Mean number of digits recalled from the constant chunk as a function of position for both immediate and delayed recall.

The overall analysis on the constant third and fifth chunks indicated a significant effect of trials, $F$ (9, 684) = 10.99, $p < .001$. The effect of position was also significant, $F$ (1, 76) = 39.93, $p < .001$, reflecting the superior recall of constant fifth chunks. In addition, the effect of delay was significant, $F$ (1, 76) = 4.00, $p < .05$, reflecting superior recall of the immediate groups. The Trials × Delay interaction was significant indicating superior improvement when recall was delayed, $F$ (9, 684) = 2.30, $p < .05$. An analysis of simple effects indicated significant improvement with both immediate, $F$ (9, 684) = 2.86, $p < .01$, and delayed recall, $F$ (9, 684) = 8.36, $p < .001$. No other interactions were significant, $F$s < 1.03.

The finding of improved recall for constant third and fifth chunks with immediate recall is at odds with the Bower and Winzenz (1969) data, and could be attributed to nonspecific practice effects. In order to partial out such effects, a difference score was calculated for constant-chunk recall, which consisted of taking the number of digits recalled in each constant chunk and subtracting from it the number of digits recalled in the same serial positions in the corresponding N-string. These scores are presented in Fig. 2. For purposes of analysis, these scores were stabilized by blocking the 10 trials for each $S$ into five blocks of 2 trials each and averaging the difference scores within each block.

An analysis of the difference scores for constant third and fifth chunks indicated a significant interaction between trials and delay; i.e., improvement was greater under conditions of delayed recall, $F$ (4, 304) = 4.30, $p < .005$. A test of simple effects indicated that the improvement across trials was not significant for immediate recall, $F$ (4, 304) = 2.24, $p > .05$, in agreement with the findings of Bower and Winzenz (1969). The improvement was significant for delayed recall, $F$ (4, 304) = 5.63, $p < .001$. No other interactions were significant, $F$s < 1. The main effect of trials was significant, $F$ (4, 304) = 2.75, $p < .05$, as was the effect of delay, $F$ (1, 76) = 4.06, $p < .05$. There was no effect of chunk position, $F < 1$.

For constant chunks in the first position the improvement across trials was significant, $F$ (4, 152) = 8.48, $p < .001$. The improvement did not interact with delay, $F < 1$. Thus, when recall is indexed by difference scores, improvement is found for any constant chunk with delayed recall, but only for a constant first chunk with immediate recall.

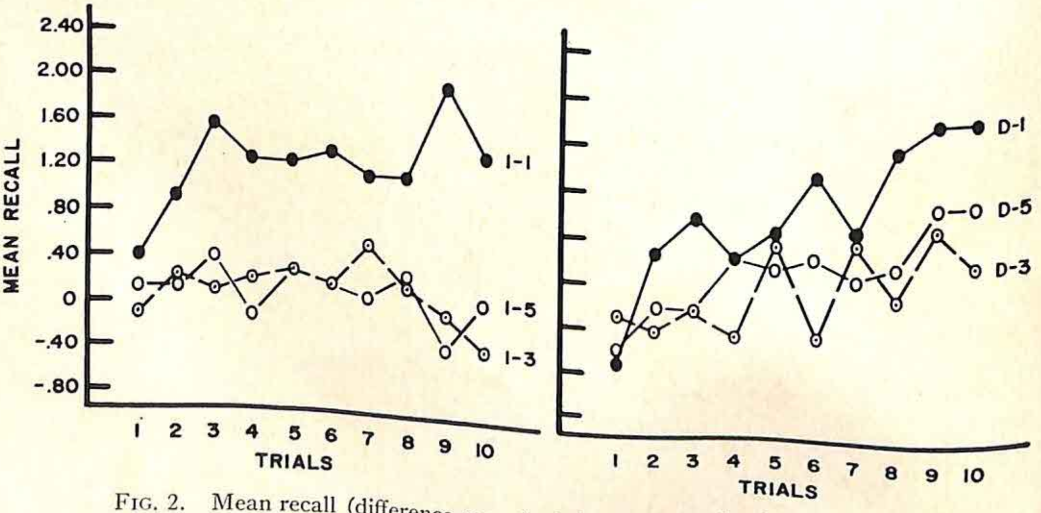


FIG. 2. Mean recall (difference scores) of the constant chunk as a function of position for both immediate and delayed recall.

### *Discussion*

The above data indicate a different pattern of improvement depending upon whether recall is immediate or delayed. One possible explanation is that the differential pattern of improvement reflects a difference in storage strategies under the two conditions. That is, under conditions of immediate recall, the memory trace which persists from trial to trial is primarily representative of only the first few chunks of the digit strings. Under conditions of delayed recall, the persisting traces tend to represent all serial positions of each digit string.

The data of Bartz (1969) have a bearing on the present data, although Bartz was exploring the repetition effect of digit strings when the entire digit string was repeated rather than a single chunk. His data indicated that the recall of digits in terminal positions of the digit strings showed greater absolute improvement when recall was delayed than for the same positions when recall was immediate. Bartz suggested that under conditions of immediate recall, the items in terminal positions of a digit string tend to be recalled primarily from STS.

An alternative possibility is that differences in immediate and delayed recall do not reflect differences in storage so much as differences in retrieval. A modification of the reallocation hypothesis could explain the data in Exp. I in terms of retrieval differences. Specifically, with delayed recall, the traces could be stored and retrieved in groups which are based upon any constant chunk, first, third, or fifth. With immediate recall, the original reallocation hypothesis would hold, and grouping of traces would be based on the first chunk. Thus, in terms of strength of storage across chunk positions, the two conditions could be equivalent, but only with delayed recall would constant third and fifth chunks improve.

## EXPERIMENTS II AND III

Experiments II and III were designed to assess whether or not the modified reallocation hypothesis might be a feasible explanation for the data obtained in Exp. I. Two types of presentation sequences were used in which four chunks of the first two (initial) strings were repeated in a third (critical) string. For the repeat condition, the critical string was a complete repeat of one of the initial strings. For the composite condition, the critical string was made up of two chunks from each of the initial strings. Experiment III contained a control condition in which the critical string and the initial strings had no chunks in common.

The modified reallocation hypothesis would predict both restricted improvement in the composite sequence and higher recall in the repeat sequence than the composite sequence. Improvement would be restricted in the composite sequence in that the critical string would have to be mapped with one previous trace or the other and thus only some of the constant chunks would be expected to improve—only the chunks in common with one of the previous strings. Recall would be superior in the repeat sequence in that all four repeated chunks could improve.

### *Method*

*Procedure and design.*—Each trial in Exp. II involved the presentation and recall of three 12-digit strings (four 3-digit chunks). Each *S* had 12 trials, with 6 being in the composite condition and 6 being in the repeat condition. Five different random orders of the trials were used. To equate for difficulty of the critical strings, two different lists were used. The critical strings in List A for the repeat condition were the same as the critical strings in List B for the composite condition, and vice versa. There were two independent groups involved in the experiment. One group recalled each string immediately after presentation and the other group recalled the strings after a delay of 15 sec. The delay procedure was the same as in Exp. I. The design was a 2 × 2 mixed design with type of recall (i.e., immediate vs. delayed) as the between-*S*s variable and type of sequence (i.e., repeat vs. composite) as the within-*S*s variable.

In Exp. III, each *S* had 18 trials instead of 12 as in Exp. II, with 6 being in the composite condition, 6 in the repeat condition, and 6 in the control condition. Seven different random orders of the trials were used. Each control trial involved the presentation of three strings, none of which had any chunks in common. To equate for difficulty of the critical strings, three different lists were used. Across the lists, the critical strings were the same for the three conditions. As in Exp. II, there were two independent groups: one immediate and one delayed.

*Materials.*—Half of the trials in the repeat condition involved repetition of the first string and half involved the repetition of the second string. For the six composite trials, there are exactly six different ways that two chunks from the first string and two from the second string can be taken to form the four chunks in the critical string. Each combination was used once. The digit sequences were generated in the same manner as in Exp. I.

*Subjects.*—Eighty-two introductory psychology students served as *S*s, 40 in Exp. II and 42 in Exp. III. In both experiments, *S*s were assigned in alternating order to one of the two groups as they came to the experiment.

### *Results and Discussion*

For the data of Exp. II (see Table 1), the critical string recall was greater than the averaged recall of the initial strings for both immediate, $F$ (1, 19) = 12.19, $p < .005$, and delayed recall, $F$ (1, 19) = 17.49, $p < .001$. These data should be regarded with caution, since, in the absence of a control condition, that improvement could be attributed to general practice effects within the presentation sequences.

The repetition effect for Exp. III (see Table 2) was assessed in terms of critical string recall. An analysis of critical string recall indicated a significant effect of sequence type for both immediate and delayed recall, $F$ (2, 40) = 3.28, $p < .05$, and $F$ (2, 40) = 7.86, $p < .005$, respectively. A Newman-Keuls test of multiple comparisons ($\alpha = .05$) indicated that for immediate recall, the only significant difference was between the control and the composite condition. For delayed recall, the control condition was inferior to both the repeat and the composite conditions.

Improvement scores were used to compare the effects of repetition between the repeat and composite sequences. The improvement score for each critical string was found by taking the number of digits recalled in the four chunks in the critical string and subtracting from it the number of digits recalled for those same chunks as they appeared in the initial strings.

TABLE 1

MEAN CORRECT DIGITS PER STRING IN EXPERIMENT II

| Sequence type | String no. | | |
|---|---|---|---|
| | First | Second | Third |
| **Immediate recall** | | | |
| Composite | 5.53 | 5.21 | 6.15 |
| Repeat | 5.45 | 5.68 | 6.06 |
| **Delayed recall** | | | |
| Composite | 3.42 | 3.40 | 4.08 |
| Repeat | 2.81 | 3.11 | 4.18 |

TABLE 2

MEAN CORRECT DIGITS PER STRING IN EXPERIMENT III

| Sequence type | String no. | | |
|---|---|---|---|
| | First | Second | Third |
| **Immediate recall** | | | |
| Composite | 6.47 | 6.16 | 6.85 |
| Repeat | 6.10 | 6.32 | 6.60 |
| Control | 6.35 | 6.10 | 5.95 |
| **Delayed recall** | | | |
| Composite | 3.88 | 4.01 | 4.75 |
| Repeat | 3.35 | 3.31 | 5.29 |
| Control | 3.17 | 3.25 | 3.34 |

The data of Exp. II did not indicate differential repetition effects for the two conditions. The mean improvement scores for the composite and repeat sequences were .71 and .79 digits for immediate recall. For delayed recall the means were .65 and 1.08. An analysis of variance indicated that the effect of delay was not reliable, $F < 1$, nor was the effect of sequence type, $F$ (1, 38) = 1.50, $p > .05$. In addition, the interaction was not significant, $F < 1$.

The data of Exp. III indicated superior improvement for the repeat sequence with delayed recall. For delayed recall the mean improvement was .84 for the composite sequence and 1.91 for the repeat sequence. For immediate recall the means were .42 and .24, respectively. An analysis of variance yielded a significant interaction between delay and type of sequence, $F$ (1, 40) = 6.12, $p < .05$. The test of multiple comparisons indicated that the difference between the sequence types in delayed recall was significant ($\alpha = .05$). The superior recall of the repeat sequence under conditions of delay is consistent with predictions made from the modified reallocation hypothesis.

Further analyses were done to assess whether or not improvement was confined to only two of the four repeated chunks in composite condition, as suggested by the modified reallocation hypothesis. For the composite condition, separate improvement scores were found for the two chunks from the first string and the two chunks from the second string. The larger score should indicate the pair of chunks that received repetition benefits and the smaller score should indicate the pair for which there was no effect. A similiar scoring procedure was used for sequences in the repeat condition, in which improvement scores were assigned to comparable pairs of chunks in the critical strings.

The difference between large and small scores should be significantly smaller in the repeat condition than the difference in the composite condition. Repetition effects should contribute to both large and small scores in the repeat condition but only to the large scores in the composite condition. The obtained mean-improvement scores were −.66 and 1.32 for the composite condition and −.31 and 1.37 for the repeat condition. A two-way analysis of variance indicated that the difference between large and small scores was not differential for the two conditions, $F < 1$. Thus, it would appear that the four chunks tend to improve in a similar fashion for both types of sequences.

In summary, these data fail to show that the improvement of composite sequences is restricted to one previous trace, as suggested by the modified reallocation hypothesis. The related prediction, that recall of repeat sequences should be superior, was borne out only in the delay conditions of Exp. III, and it would seem that that superiority would have to be accounted for in terms of some hypothesis other than reallocation.

The failure to find superior recall of the repeat sequences in immediate recall could possibly be due to the small repetition effects obtained with the present procedure. The same explanation could apply to the delayed recall data of Exp. II; however, without the appropriate control condition for that data, it is impossible to make any valid inferences.

The data of Exp. II and III offer further support for the finding in Exp. I that improvement is restricted to the first few chunks in immediate recall. Separate improvement scores (maximum = 3) for each chunk position were found for the data of Exp. II and III, and are given in Table 3. In general, improvement is confined to the first few chunk positions when recall is immediate and more spread out with delayed. The score for the first chunk in the delayed-composite condition (Exp. II) was insignificantly negative; i.e., recall was inferior when that chunk was repeated. There is no obvious reason why that should have occurred and it was not replicated in Exp. III.

TABLE 3

IMPROVEMENT SCORES FOR EACH CHUNK POSITION

| Sequence type | Chunk position | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Exp. II | | | | Exp. III | | | |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Immediate | | | | | | | | |
| Composite | .29* | .28* | .08 | .05 | .15 | .17 | .13 | −.03 |
| Repeat | .48** | .29* | .00 | .03 | .19* | .12 | −.13 | .07 |
| Delayed | | | | | | | | |
| Composite | −.23* | .25* | .35** | .28** | .20* | .20* | .21* | .23 |
| Repeat | .28** | .17 | .34* | .30* | .42** | .51** | .38* | .60** |

* $p < .05$.
** $p < .01$ or beyond.

## General Discussion

These data appear to rule out the reallocation hypothesis, in either its original or modified versions, for delayed recall. One hypothesis that might adequately account for the delayed recall data in the present set of experiments is the bin hypothesis (Bower & Winzenz, 1969). According to the bin hypothesis, memory consists of a small set of storage slots or bins. The first chunk of every digit string becomes associated with the first bin, the second with the second bin, and so on.

The bin hypothesis would predict improvement for a repeated chunk in any position, which is consistent with the delayed recall data of Exp. I. In addition, it would predict that in the composite sequences of Exp. II and III, improvement could be found for repeated chunks from two different previous traces. However, one aspect of the delayed recall data that is not consistent with the bin hypothesis is the superior recall of repeat sequences over the composite sequences in Exp. III.

It is possible that the advantage of the repeat condition is due to repetition of inter-chunk information provided in that condition. Bower and Winzenz (1969, Exp. II) demonstrated that inter-chunk learning takes place when an entire digit string is repeated. In the present studies, when the four chunks are repeated in the repeat condition, the same inter-chunk information is repeated. When the four chunks are repeated in the composite condition, the inter-chunk relationships change from what they were in the initial strings.

It is not immediately obvious how one would incorporate the notion of between-chunk learning into the framework of the bin hypothesis. One possibility is to hypothesize stored associations between successive chunks in adjacent bins. Another possibility is that higher order codes are formed to retrieve information from successive groups of bins (e.g., Johnson, 1968). It seems clear, however, that some notion of between-chunk learning, together with the bin hypothesis, could account for the delayed recall data.

If the same hypothesis is to account for both immediate and delayed recall, it might be argued that the bin hypothesis also holds for immediate recall. Different repetition effects in the two conditions might be explained in terms of more extensive LTS storage in delayed recall. There are immediate recall data, however, obtained by Bower and Winzenz (1969), that are inconsistent with the bin hypothesis and which cannot readily be accounted for in terms of lack of LTS strength (viz., Exp. VI and VII).

Thus, it is possible that one explanation pertains to immediate recall and another to delayed recall. The reallocation hypothesis appears to explain, quite well, the immediate recall data of Bower and Winzenz (1969) and that in the present experiments.

## REFERENCES

Bartz, W. H. Repetition and the memory stores. *Journal of Experimental Psychology*, 1969, **80**, 33–38.

Bower, G. H., & Winzenz, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, 1969, **80** (2, Pt. 2).

Johnson, N. F. Sequential verbal behavior. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs: Prentice-Hall, 1968.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

(Received December 10, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 417–424

# EFFECT OF RETENTION INTERVAL, RETROACTIVE INHIBITION, AND PROACTIVE INHIBITION ON MEDIATING ASSOCIATIONS[1]

ALLEN R. DOBBS[2]

*University of Iowa*

Traditional proactive inhibition (PI), retroactive inhibition (RI), and Rest Control retention paradigms were combined with A-B, B-C, A-C mediation, and A-B, D-C, A-C control paradigms to form 12 groups ($n = 20$ per group). A 10-min. or 24-hr. retention interval was imposed immediately after B-C (D-C) learning. In accordance with a mediational (chaining) analysis, both A-B recall taken immediately prior to A-C learning and the degree of mediation paradigm superiority during A-C indicated: (*a*) forgetting over a 24-hr. interval (*b*) greater RI than PI on immediate test, and (*c*) decreasing RI over the interval (only a trend for increasing PI). The results are counter to expectations from pseudomediational and direct-association interpretations.

There have been numerous demonstrations that A-C performance in the context of the A-B, B-C, A-C mediational (M) transfer paradigm yields performance superior to that obtained on the same A-C list but in the context of the A-B, D-C, A-C control (C) paradigm. The theoretical interpretation has usually been in terms of the ability of *S*s receiving the M paradigm to use the prior-learned A-B and B-C associations to mediate A-C during Stage 3 learning. This chaining interpretation leads to the prediction that any variable which alters the strength of either or both of the mediating A-B and B-C associations should likewise produce a change in the level of superiority of the M over the C paradigm. Although the manipulation of acquisition variables has yielded some support for this prediction (e.g., Schulz, Liston, & Weaver, 1968; Schulz, Weaver, & Dobbs, 1968), investigations in the area of retention have been few and conflicting. The major premise of the present study was that the known effects of proactive inhibition (PI), retroactive inhibition (RI), and retention interval obtained with "traditional" retention paradigms could be generalized to the case of the *mediating* associations assumed to be employed by *S*s receiving an M paradigm. On the basis of the premise one would expect combined retention and mediation paradigms to produce relative magnitudes of mediation effects indicating: (*a*) forgetting over an interval, (*b*) greater RI than PI with an immediate test, but (*c*) decreasing RI and increasing PI over a retention interval. Contrarily, no differential effect of the various retention variables would be expected for the A-C performance of the C paradigms when appropriate materials are employed (cf. Schulz, Weaver, & Ginsberg, 1965).

Not only is there little evidence by means of which these predictions can be evaluated, but the evidence which is available is conflicting. For example, Murdock (1952) and Adams (1964) have reported evidence supporting the predictions in that they found (respectively) retroactively and proactively interfering associations reduced mediated performance. On the other hand, Peters (1935) obtained a trend for better mediation after 24 hr. than immediately, and Horton and Wiley (1967) found no decrement in mediation even after a 10-day retention interval, even though A-B recall taken *after* A-C learning showed larger decrements with longer retention intervals. However, none of these studies evaluated the state of the interfered with or to-be-remembered association *prior* to the time it was employed mediationally. Such a measure is requisite in that corroboration for the chaining analysis must be assessed in terms of parallel effects of the retention variables on the association's strength when directly recalled and employed mediationally.

---

[1] This report is based on a dissertation submitted to the University of Iowa graduate school in partial fulfillment of the requirements for the PhD degree. The research was supported in part by National Science Foundation Grant GB 2835 awarded to Rudolph W. Schulz. Sincere appreciation for the many invaluable consultations is expressed to Rudolph Schulz, who served as chairman of the author's doctoral committee.

[2] Requests for reprints should be sent to Allen R. Dobbs, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

The approach taken herein was to combine the "traditional" PI and RI paradigms with the M and C paradigms such that an interfering list would produce PI or RI for the A-B association. Rest-control groups also received the M and C paradigms, but no additional interfering list was given. Two retention intervals were employed with an A-B recall trial administered immediately before and after A-C learning.

## METHOD

*Design.*—The experiment consisted of a 2 × 2 × 3 factorial design. The respective sources of classification were, paradigm (Mediation or Control), retention interval (10 min. or 24 hr.), and subset (PI, RI, or Rest Control). Eight of the 12 groups received an interfering A-X list. In order to establish a PI paradigm, A-X learning preceded List A-B for the two M and two C groups of the PI subset. The interpolation of A-X learning between Lists A-B and B-C (D-C) formed RI paradigms for the other 4 groups receiving an interfering list. The Rest-Control subset was not given an interfering list and provided control groups for the PI and RI subsets. The first A-B recall, A-B Recall-1, was given immediately prior to A-C learning and enabled an assessment of the state of the "direct" A-B association. Assuming a constant level of B-C strength for all M groups, Stage 3 (A-C) performance provided for an evaluation of A-B strength when this association was used mediationally. Hence, if PI, RI, and forgetting are demonstrated in A-B Recall-1, the major premise of the present study will be supported to the extent that A-C mediation effects for the PI, RI, and Rest-Control conditions are found to reflect performance obtained on the first recall trial. The second A-B Recall trial, A-B Recall-2, was given immediately after A-C learning and when compared with A-B Recall-1 provided for an assessment of the extent to which an A-B recall trial given *after* Stage 3 indicates the state of this association prior to A-C.

*Subjects.*—The *S*s were 240 University of Iowa undergraduates whose experimental participation was in fulfillment of a course requirement. Ten males and 10 females were assigned randomly to each group, with the restriction that each group must have $n - 1$ *S*s prior to the assignment of the $n$th *S* to any group. Since the 24-hr. conditions necessitated *S*s' participating in the experiment on 2 successive days, all *S*s were required to sign up for 2 days to avoid *S* selection, but those in the 10-min. groups were dismissed after the first session.

*Materials.*—Each list was comprised of six pairs of items. The A stimulus terms were paralogs with a mean association value of 38.7 (Taylor & Kimble, 1967). The formal similarity of those terms was minimal. The B, D, and X terms were AA nouns (Thorndike & Lorge) with minimal associative overlap. The C terms, which were CVC trigrams of low formal similarity, had an average association value of .42 (Noble, 1961). Four randomizations of each list were used to prevent serial learning. These particular classes of materials were employed because they, in conjunction with the procedures outlined below, produce a situation not subject to pseudomediation interpretations (Earhard & Mandler, 1965, p. 365).

*Procedure.*—The general methods employed were similar to those described by Schulz et al. (1965) and Schulz and Weaver (1968). All lists were learned by the study–test method. The study interval was 2 sec. for all lists, and the test interval was 2 sec. for the A-B and A-X lists, but 3 sec. for the B-C (D-C) and A-C test trials. With the A-B and A-X lists, *S*s recalled the responses on test trials. The test trials for the B-C (D-C) and A-C lists were of the multiple-choice variety where each stimulus term was presented with four alternatives, the correct response and three other list-member response terms.

In order to equate the total time between the end of A-B learning, which marked the beginning of the retention interval, and A-B Recall-1, the PI and Rest-Control subsets received a filled rest interval equal to the time required to administer the A-X list to the RI subset (2 min.). This rest interval consisted of working alternating addition and subtraction problems with instructions emphasizing speed and accuracy.

At the end of B-C (D-C) learning, *S*s serving under the 24-hr. conditions were dismissed, having already been scheduled to return in 24 hr. Since the total time necessary to administer the RI A-X list (or the rest interval task) and B-C (or D-C) was approximately 10 min., this was the minimum retention interval possible for the recall of A-B. At the end of the retention interval (immediately following B-C or D-C for the 10-min. groups, or 24 hr. later for the 24-hr. groups), the A-B Recall-1 trial was given.

In order to assess the major premise of the present study, it was desirable to equate as nearly as possible the conditions under which the effects of the retention variables were evaluated when the A-B association was directly recalled and when it was employed mediationally. When the association is used mediationally, *S*s are assumed to chain from A to a covert *recall* of B, and from B to C, all within the 3-sec. test interval. Consequently, when the association is used mediationally, the availability of B is important and there are time restrictions placed on its recall. For these reasons, A-B Recall-1 and -2, in which the state of the "direct" A-B association was evaluated, were paced-recall trials (2-sec. rate) as opposed to an unpaced modified-modified free recall (MMFR) task. Since it was the manipulation of the covert recall of the A-B portion of the mediational chain that was of interest, an attempt was made to minimize the influence of variables which might alter the utilization of the B-C link. This was accomplished by employing a multiple-choice recognition procedure as it was thought that this would reduce or eliminate any confounding influence of differential C-term availability or B-C associative forgetting as a result of the 24-hr. interval. It should be noted that the comparison of the A-B association is still in terms of recall in both the "direct" and mediational situations. Further, it can be seen that the C groups are the appropriate controls in that it is only to the extent that the M groups can recall A-B (*albeit* covertly) that their performance can surpass that of the C conditions.

*Degree of learning.*—The criterion for the A-B, B-C, and the D-C lists was two trials, not necessarily consecutive, on which all items were correctly given. A-C learning consisted of 10 alternating study and test trials. By means of a pilot study, the number of trials necessary to acheive a criterion of 5/6 correct was determined when A-X was the first list (as in the case of PI) and when it was preceded by A-B (as in the case of RI). The numbers of trials were found to be 8 and 5, respectively; hence all PI groups received 8 trials of A-X learning, while 5 A-X trials were administered to the RI groups.

## RESULTS

*Stage 1.*—The overall trials-to-criterion mean for A-B learning was 8.25, with a range extending from 7.05 to 9.55. None of the factors or interactions were reliable at the .05 level. The comparable A-B performance of the PI groups which received A-B and A-X, and the remaining groups which received A-B as a first list, indicates the positive learning-to-learn and warm-up factors counteracted the negative transfer from A-X to A-B.

*Stage 2.*—It will be recalled the paradigms differed in this stage, the M and C groups learning B-C and D-C lists, respectively. The trials-to-criterion means ranged from 6.15 to 8.05 for B-C with an overall mean of 6.92, and from 5.35 to 6.65 for D-C with an overall mean of 5.99. The M group mean for B-C was found to exceed that of the C group for List D-C in each pair of M and C groups. The analysis showed the trend for the superiority of the C over the M groups to be reliable, $F\ (1, 228) = 3.98$, $p < .05$. All other $F$s were less than 1.00. The finding of a D-C superiority in Stage 2 is in accord with previous studies (e.g., Schulz et al., 1965) using similar materials. Separate intraparadigm comparisons for the M and C groups showed that the differences among the six B-C and six D-C means were no larger than would be expected from random sampling.

*A-X learning.*—The mean total numbers of correct responses over the eight trials of List A-X given by the PI subset groups ranged from 29.55 to 34.65. Only the retention interval variable approached statistical reliability, $F\ (1, 76) = 3.28$, $.05 < p < .10$. The means for this comparison are 31.22 for the 10-min. interval and 34.52 for the 24-hr. interval groups. This result implies that the degree of learning attained by the PI groups in the 24-hr. condition may have been slightly higher than that attained by the 10-min. PI groups. The mean total numbers of correct responses obtained by the RI subset groups over the 5 A-X trials ranged from 13.45 to 15.70. None of the $F$s approached the .05 level of significance, indicating the RI groups attained comparable levels of A-X learning.

An analysis performed on the mean terminal levels of learning for the PI and RI subsets (Trials 8 and 5, respectively) indicates the PI groups attained a reliably higher level of performance on the A-X list than did the RI groups, $F\ (1, 159) = 64.53$, $p < .01$. The means for this analysis were 5.18 and 4.58 for the PI and RI combined groups, respectively. Hence it appears that the attempted equation for the degree of A-X learning by the pilot study was not entirely successful. The implications of this will be taken up in the discussion section.

*A-B Recall-1.*—It will be recalled that the lists given up to, and including, A-B Recall-1 essentially conformed to the traditional PI, RI, and Rest-Control paradigms. The difference was that a B-C or D-C list was interpolated between the end of the second list (or first list in the case of the Rest-Control groups) and the recall trial. As stated earlier, this recall trial was given in an effort to ascertain the state of the "direct" A-B association just prior to

TABLE 1

Mean Numbers of Correct Responses on A-B Recall-1 and A-B Recall-2, and Mean Total Numbers of Correct Responses over 10 A-C Trials for Each Group

| Task | PI | | | | RI | | | | Rest Control | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M-10 | C-10 | M-24 | C-24 | M-10 | C-10 | M-24 | C-24 | M-10 | C-10 | M-24 | C-24 |
| A-B Recall-1 | 3.30 | 3.80 | 1.45 | 2.20 | 1.90 | 2.60 | 2.75 | 4.00 | 3.85 | 4.90 | 3.15 | 3.75 |
| A-C | 39.80 | 32.20 | 30.10 | 31.65 | 37.70 | 33.40 | 35.70 | 28.75 | 44.85 | 31.40 | 35.20 | 28.95 |
| A-B Recall-2 | 3.30 | 3.50 | 1.75 | 2.40 | 2.20 | 2.15 | 3.00 | 2.85 | 4.50 | 4.55 | 3.70 | 3.65 |

the A-C test list where it could be used as a mediating association by the M groups. The following analyses provided a determination of the effectiveness of the present paradigms in producing the effects previously obtained for PI, RI, and Rest-Control paradigms.

The mean numbers of A-B items correctly recalled in A-B Recall-1 shown in the first row of Table 1. As can be seen from this table, *S*s in each of the six C-paradigm groups recalled a greater number of items than did *S*s in the paired M groups. Statistically, this difference is a reliable one, $F\ (1,\ 228) = 13.54$, $p < .01$. This finding is contrary to previous findings from the Iowa Laboratory (e.g., Schulz, Weaver, & Dobbs, 1968). The PI groups recalled a greater number of items than the RI groups at the 10-min. interval; however, this trend was reversed after 24 hr. The Retention Interval × Subset interaction indicates this trend to be reliable, $F\ (2,\ 228) = 16.17$, $p < .01$. None of the interactions involving the paradigm variable gave *F*s greater than unity. Because of the Retention Interval × Subset interaction, subset differences were evaluated independently at each retention interval. Since the paradigm variable did not interact with either of the other two factors, the respective M and C groups were averaged. The analysis of the combined M and C groups showed the subset variable to have yielded reliable differences at both intervals, $F\ (2,\ 228) = 24.26$, $p < .01$ and $F\ (2,\ 228) = 13.95$, $p < .01$, for the 10-min. and 24-hr. intervals, respectively.

Comparisons from Table 1 show a trend for both increasing PI and decreasing RI relative to the Rest-Control conditions. However, the interaction which tested the trend for increasing PI only approached statistical significance, $F\ (1,\ 152) = 2.41$, $.05 < p < .10$. Contrarily, the interaction $F$ for decreasing RI achieved reliability, $F\ (1,\ 152) = 17.92$, $p < .01$. Further, a one-tailed $t$ test evaluating the reliability of the observed recall superiority of the combined RI 24 groups over that of the RI 10 groups showed the trend to be reliable, $t\ (228) = 1.75$, $p < .05$. This indicates that reliable absolute spontaneous recovery of the first-list associations was demonstrated for the RI paradigm.

*Stage 3.*—The objectives of the analysis of Stage 3 data were to determine the comparability of the C groups on List A-C, the reliability of the mediation effects for each of the M-C pairs, and to assess the effects of the retention paradigms at the two retention intervals on the magnitude of the mediation effects. Since a graph of Stage 3 acquisition functions showed there was no interaction save a slight tendency for the curves of the various groups to converge toward a common asymptote, the mean numbers correct over the 10 A-C trials were used in all Stage 3 analyses.

The six C groups all performed at about the same level on List A-C. The means ranged from 28.75 to 33.40 with an overall mean of 31.07, and statistically the C group differences can be considered as chance deviations.

The mediation effect was evaluated separately for each of the respective M and C groups, using an overall error term. In accordance with the trends shown in the second row of Table 1, all M minus *C* differences were found to be reliable at the .05 level, or less, except that for the PI 24 groups.

Differences between M 10 and C 10, and M 24 and C 24 show a trend for a decreased mediational effect after 24 hr. for the Rest-Control groups. This decrease, which statistically is a Paradigm × Retention Interval interaction, was found to only approach statistical reliability, $F$ (1, 76) = 3.25, $.05 < p < .10$. However, a test contrasting only Cond. M 10 with M 24 indicated poorer performance was obtained for the Rest-Control M paradigm with a 24-hr. retention interval, $t$ (228) = 3.17, $p < .01$.

There is a slight trend for the PI subset to have resulted in a greater decrement over the 24-hr. interval than resulted from the forgetting of the Rest-Control groups. A test of this interaction yielded an $F$ of less than unity. Hence the predicted increase in PI was not obtained for either the A-B Recall-1 or the A-C data.

A comparison between the 10-min. and 24-hr. RI groups indicates a strong trend for the decrement in the mediation effect under RI conditions to be greater at the 10-min. interval than after 24 hr. relative to the Rest-Control groups, although the interaction did not achieve the .05 level of statistical significance, $F$ (1, 152) = 2.65. However, since the parallel C groups have a severe dampening effect, from a statistical standpoint, on the likelihood that the interaction in question would achieve significance, a different approach to the test of what graphically is clearly a trend toward an interaction was taken which eliminated this attenuation. That is, the mean of the respective control groups was subtracted from the score of each of the RI-M and M group *S*s. A 2 × 2 analysis of variance (Paradigm × Retention Interval) conducted on these scores produced a reliable interaction, $F$ (1, 76) = 5.55, $p < .025$, indicating spontaneous recovery of the A-B association over the retention interval was obtained for the RI condition.

*Analysis of A-B Recall-2.*—The analysis of the second A-B recall trial had as its objective an evaluation of the state of the A-B association after List A-C. Since only the A-C list intervened between A-B Recall-1 and -2, any differences between these recall measures should reflect processes occurring during A-C learning. The mediational analysis assumes that the A-B association is used by the M groups because, in conjunction with the B-C association established during Stage 2, it leads to the correct response (C) in List A-C. Hence, the A-B association should be maintained or strengthened in the M groups. Contrarily, the A-B association does not lead to the correct response with the C paradigm and hence, it may be weakened or lost.

The mean numbers of correct recalls for each of the groups are shown in the last row of Table 1. As can be seen, the reliable M versus C group differences obtained in A-B Recall-1 have disappeared for all pairs except for the PI-M 24 versus PI-C 24 group comparison, which, it will be recalled, was also the only comparison which did not show a mediational effect. The paradigm variable yielded an $F$ of less than 1.00 indicating the reliable M versus C group differences found for A-B Recall-1 were no longer present in A-B Recall-2. However, the Retention Interval × Subsets interaction found in A-B Recall-1 was also reliable for the A-B Recall-2 data, $F$ (2, 228) = 9.19, $p < .01$.

The A-B Recall-2 data were also analyzed to test for decreasing and increasing recall for the PI and RI groups, respectively, relative to the Rest Controls. The test for an increasing decrement in recall for the PI groups did not approach statistical reliability, $F$ (1, 152) < 1.00. As in A-B Recall-1, the test for increasing recall after 24 hr. for the RI group relative to the Rest Controls was reliable, $F$ (1, 152) = 11.65, $p < .01$.

## DISCUSSION

A comparison of the trends displayed in the first and second rows of Table 1 shows the striking similarity of the effects of the retention variables on the A-B association when it was recalled overtly (A-B Recall-1) and when it was employed mediationally (List A-C). The

Rest-Control groups showed a decrease in recall of the A-B association in A-B Recall-1 after a 24-hr. interval as compared to a 10-min. retention interval. Similarily a direct comparison of Rest-Control M groups also shows a decrement in performance at the longer retention interval when the A-B association was employed mediationally in A-C. Both RI and PI were obtained at the 10-min. interval when the A-B association was recalled and when used mediationally. Decreasing RI was obtained in both the recall- and mediation-test phases. Absolute spontaneous recovery was obtained in A-B Recall-1, although comparisons gave evidence for only relative spontaneous recovery during the mediation-test phase. Reliable increasing PI was not obtained in A-B Recall-1, and as would be predicted on the basis of this, the mediation data also revealed only a slight unreliable trend.

The crossover of the RI and PI groups from the 10-min. to the 24-hr. interval in A-B Recall-1 was an unexpected finding. With an equal degree of learning on the interfering list, only a convergence would be expected on the basis of previous research. However, despite the attempted equation by means of a pilot study, there were reliable differences in the degree of A-X learning achieved by the RI and PI subsets, with the PI groups attaining the higher level. Further, the difference is probably underestimated by the response measure, as the majority of the PI *S*s attained all items correct on Trial 8. The effect of a greater amount of PI relative to RI as obtained in the present study would be to lower the converging PI function, thereby increasing the probability of a crossover between the PI and RI recall functions. Hence, the crossover found herein is not surprising, since it probably resulted from the unsuccessful equation of A-X learning for the PI and RI subsets. Nevertheless, in accordance with expectations, the crossover obtained in A-B Recall-1 was also evidenced in the Stage 3 mediated performance.

It should be noted that A-B Recall-1 preceded the mediation-test phase for all groups and this may be thought to restrict the generality of the results obtained during A-C learning. However, it is doubtful that this is the case. Since *S*s recalled A-B during A-B Recall-1 and received no feedback, and then had to recall A-B in order to mediate A-C, the question of what would have happened in A-C had the first recall trial not been given can be estimated from our knowledge of the effect of an A-B recall trial on future A-B recalls. Previous research indicates the effect of a "nonreinforced" A-B recall trial is to but slightly enhance the potential that B would be better recalled on a future occasion, in the case where there was (Richardson & Gropper, 1964) and was not (Goss, Nodine, Gregory, Taub, & Kennedy, 1962), an interpolated list, and after a 24-hr. retention interval (Wollen, 1962). This implies that in the present study any *differences* in the strength of A-B produced by the retention variables should have been left the same, although the absolute levels of recall for all groups may increase. There should have been no effect of this recall trial on the A-C performance of the C group. Consequently, it would seem that the generality of the present A-C findings are limited by the inclusion of A-B Recall-1 only in that one might expect *larger* differences between the M groups to be obtained when this recall trial is omitted.

The superior recall of the C over their respective M groups in A-B Recall-1 is not in accord with previous mediational investigations from the Iowa laboratories. For example, Schulz et al. (1968) found B-C versus D-C learning interpolated between A-B learning and recall did not have a differential effect on the recall of A-B, even after 22 B-C or D-C trials. In examining the procedures used in previous investigations and those of the present study, none of the procedural differences would be expected to result in this effect. Consequently, the superior recall of the A-B association for C groups during A-B Recall-1 necessitates a discussion of the A-C data in terms of pseudomediation. If it appears a pseudomediational interpretation can account for the Stage 3 data, then no conclusions can be drawn from the present research with respect to the effect of retention variables on mediational events.

The pseudomediation interpretation (Mandler & Earhard, 1964) analyzes the M and C paradigms as essentially A-B, A-C negative transfer paradigms, with a B-C or D-C interpolated list. The superior performance of the M paradigm during A-C is assumed to be due to the unlearning of the interfering A-B association during B-C (D-C). With this interpretation, two prior-learned lists should be more interfering than one, and the order in which these interfering lists (A-B and A-X) are learned should not lead to differential predictions since B and X function merely as interfering associations. As was reported earlier, no differential negative transfer for one versus two prior-learned "interfering" lists was obtained in the present study for the C groups. Further, the order in which the A-B and A-X lists were learned (i.e., PI vs. RI) *did* have differential effects upon the M groups, and these effects changed from the 10-min. interval to the 24-hr. interval in accordance with a mediational analysis, but contrary to expectations from a pseudomediation analysis. Hence, the present results show that predictions in terms of overall A-C performance indicate the pseudomediational analysis must be rejected. In addition, a pseudomediation analysis would predict a decrement in performance on A-B Recall-2 compared to A-B Recall-1 for both M and C paradigms as a result of learning an interfering A-C association. Contrarily, a mediational analysis would predict better recall of A-B after A-C learning for all M groups showing a mediation effect. This increment in recall is assumed to result from the utilization of the A-B association ($A \rightarrow B \rightarrow C$). In terms of the present study, this means that the differences between the respective M and C groups should have been reduced, eliminated or reversed for all comparisons except PI-M 24 and PI-C 24 where no mediation effect was obtained. The findings are contrary to a pseudomediation interpretation but confirmatory with respect to a mediational analysis. Finally, an item analysis showed that A-C performance was positively correlated with the occurrence of a correct recall of A-B for the M groups, but no consistent relationship between A-B Recall and A-C performance was obtained for the C paradigm as would be expected from a mediational interpretation. Since both in terms of overall A-C performance and internal item analyses the data are inconsistent with predictions from a pseudomediational interpretation, this interpretation can be conclusively rejected with respect to the present data.

A different mediational analysis has been proposed by Horton and Kjeldergaard (1961) and Schwenn and Underwood (1965), and more recently by Kjeldergaard (1968) and Peterson and Koltnow (1968). According to this analysis, mediation and the formation of the A-C association occur during B-C learning. Here it is assumed that A is elicited by B through a backward association and the A-C association is formed through the contiguity of the A and C terms ($B^{\nearrow} - {}^{\searrow}C$). However, this analysis cannot account for the increased mediational effect obtained after 24 hr. with the RI relative to the Rest-Control paradigm. Specifically, the A-X list which followed A-B learning should have decreased the backward elicitation of A during B-C, thus allowing for little formation of the A-C association. The prediction would then be for lower A-C performance for the RI-M 10 group as compared to the 10-min. Rest Control as was obtained. However, the 24-hr. interval should serve to reduce the mediation effect for both the RI and Rest-Control groups under the second-stage hypothesis, as forgetting of the A-C association acquired during B-C learning should occur. Instead, the results showed a relative increment in the mediation effect for the RI 24 condition, indicating spontaneous recovery of the first-list association facilitated Stage 3 chaining.

The results of the present study are not in accord with the findings obtained in previous investigations concerned with the retention of a mediating association. Whereas Peters (1935) found no mediation with an immediate test, but a small trend for mediation after 24 hr., the present study found strong mediation at both intervals for those groups not receiving a list specifically designed to interfere with one of the mediating associations. In addition, the mediation was *less* after 24 hr. This latter finding is also clearly contrary to the findings of Horton and Wiley (1967), who found no decrement in the mediational effect even after a 10-day retention interval. Although the discrepancy between the present findings and those of Peters may, in part, be due to the different mediational paradigms employed, paradigm differences can not account for the differential results obtained in the present study and that of Horton and Wiley. This latter study did employ slightly higher degrees of Stage 1 and 2 learning, and a recall procedure during A-C learning as opposed to the recognition task used in the present study. However, there appears to be no reason to expect these differences to have produced the discrepant results. Further, since the present results indicate that an A-B recall measure taken after A-C does adequately reflect the state of the A-B association prior to A-C, the Horton and Wiley study apparently obtained strong A-C performance at all retention intervals even though the A-B association was decreasing in strength as measured by the recall of A-B taken after A-C. At present, it appears no reconciliation of the results reported herein with those obtained by Horton and Wiley can be made.

## REFERENCES

Adams, J. A. Mediation and forgetting. *Psychological Reports*, 1964, **14**, 591–600.

Earhard, B., & Mandler, G. Mediated associations: Paradigms, controls, and mechanisms. *Canadian Journal of Psychology*, 1965, **19**, 346–378.

Goss, A. E., Nodine, C. F., Gregory, B. N., Taub, H. A., & Kennedy, K. E. Stimulus characteristics and percentage of occurrence of response members in paired-associates learning. *Psychological Monographs*, 1962, **76** (12, Whole No. 531).

Horton, D. C., & Kjeldergaard, P. M. An experimental analysis of associative factors in mediated generalization. *Psychological Monographs*. 1961, **75** (11, Whole No. 515).

Horton, D. L., & Wiley, R. F. The effect of mediation on the retention and strength of previously formed associations. *Journal of Verbal Learning and Verbal Behavior*, 1972, **6**, 36–41.

Kjeldergaard, P. M. Transfer and mediation in verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Mandler, G., & Earhard, B. Pseudomediation: Is chaining an artifact? *Psychonomic Science*, 1964, **1**, 247–248.

Murdock, B. B., Jr. The effect of failure and retroactive inhibition on mediated generalization. *Journal of Experimental Psychology*, 1952, **44**, 156–164.

Noble, C. E. Measurements of association value (*a*), rated associations (*a′*), and scaled meaningfulness (*m′*) for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, **8**, 487–521.

Peters, H. N. Mediate association. *Journal of Experimental Psychology*, 1935, **18**, 20–48.

Peterson, J. J., & Koltnow, J. Retention of the acquisition pairs in a mediation paradigm before and after the test trial. *Journal of Experimental Psychology*, 1968, **76** (3, Pt. 2).

Richardson, J., & Gropper, M. S. Learning during recall trials. *Psychological Reports*, 1964, **15**, 551–560.

Schulz, R. W., Liston, J. R., & Weaver, G. E. The A–B, B–C, A–C paradigm: Recall of A–B following A–C learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 602–607.

Schulz, R. W., & Weaver, G. E. The A-B, B-C, A-C, mediation paradigm: The effects of variation in study- and test-interval lengths and strength of A-B or B-C. *Journal of Experimental Psychology*, 1968, **76**, 303–311.

Schulz, R. W., Weaver, G. E., & Dobbs, A. R. The A-B, B-C, A-C chaining paradigm: The effect of A-C performance of varying numbers of B-C trials. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1968.

Schulz, R. W., Weaver, G. E., & Ginsberg, S. Mediation with pseudomediation controlled: Chaining is not an artifact! *Psychonomic Science*, 1965, **2**, 169–170.

Schwenn. E. & Underwood, B. J. Simulated similarity and mediation time in transfer. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 476–483.

Taylor, J. D., & Kimble, G. A. The association value of 320 selected words and paralogs. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 744–752.

Wollen, K. A. One-trial versus incremental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 14–21.

(Received November, 1 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 425–427

## SHORT REPORTS

# DETECTION OF REDUNDANCY IN MULTIPLE CUE PROBABILITY TASKS[1]

BRIAN A. KNOWLES,[2] KENNETH R. HAMMOND, THOMAS R. STEWART

*University of Colorado*

AND DAVID A. SUMMERS

*University of Kansas*

Thirty-six *S*s completed a two-cue probability learning task in which both the sign (positive or negative) and magnitude (high or low) of stimulus redundancy (cue intercorrelation) were varied. Subsequently, *S*s judged each of a set of cue combinations as being likely or unlikely to occur. For each *S*, the cue intercorrelation and standard error of estimate were determined from the combinations judged likely to occur. Learning of the cue intercorrelations did occur, although the accuracy of *S*s' estimates varied with both sign and magnitude of the intercorrelation. Previous results from cue-criterion learning studies were thus found to generalize to the learning of cue intercorrelations.

Recently, Knowles, Hammond, Stewart, and Summers (1971) found that *S*s' performance in a probabilistic inference task varied with both the *sign* (positive or negative) and the *magnitude* of intercorrelation (redundancy) between the cues. Specifically, in a two-cue probability learning task, learning was most rapid when there was a high positive correlation and least rapid when there was a high negative correlation between the cues. However, it was not possible to determine from their data whether that result arose from differential difficulty in the *detection* of the intercorrelations, or indeed whether *S*s detected the intercorrelation at all.

The present experiment was designed to address the question of *S*s' detection of cue intercorrelations. This was done by investigating the extent of *S*s' learning of the relations between cues while their attention was focused on learning the relations between the cues and the criterion. The same inference task used by Knowles et al. (1971) was used in the present study.

*Method.*—The *S*s were 36 introductory psychology students who participated in order to fulfill a course requirement. They were randomly assigned to one of four training conditions, 9 *S*s per condition. In each training condition, *S*s completed a 50-trial task in which they were instructed to learn to make inferences from two probabilistic cues. The task was a simulated medical diagnostic task requiring *S*s to estimate a criterion variable on the basis of two supposed medical cues. Four different tasks were constructed, one for each condition, where each task involved a different cue intercorrelation. The criterion values (correct answers) for all four tasks were linearly related to the cue values, with the weight for Cue 2 being approximately twice the weight for Cue 1. The linear relation was not perfect since random error was added to the criterion so that the squared multiple correlation between the two cues and the criterion value was .84. All task characteristics for each condition (see Table 1) remained approximately constant for each block of 10 trials. The tasks and instructions are described in more detail in Knowles et al. (1971). At no time were *S*s told anything about the correlation between the cues nor were they instructed to attend to this aspect of the task.

Following completion of the learning task, all *S*s were given 50 test trials similar in appearance to the training trials. The cues in these test trials were uncorrelated for each block of 10 trials and uniformly distributed throughout their range. In the test phase, *S*s were asked to judge, on the basis of their training experience, which of the test trials were *likely* and which were *unlikely* to occur. They were told to put as many instances as they liked into either category. If *S*s had (incidentally) learned the correlation between the two cues in the training stage, they would find some of the test trials to be inconsistent with their experience. For example, if the cues were intercorrelated .8, then a trial in which one cue had a very high value and the other had a very low value would be unlikely to occur.

TABLE 1

TASK CHARACTERISTICS

| Redundancy condition ($r_{12}$) | Characteristic | | | | |
|---|---|---|---|---|---|
| | $r_1$ | $r_2$ | $b_1$ | $b_2$ | $R_e^2$ |
| .8 | .832 | .896 | .320 | .640 | .84 |
| .4 | .642 | .856 | .357 | .714 | .84 |
| −.8 | −.410 | .820 | .683 | 1.366 | .84 |
| −.4 | .099 | .795 | .497 | .994 | .84 |

Note.—$r_{12}$ is the correlation between Cue 1 (*REL. B. V.*) and Cue 2 (NIPROG.); $r_1$ is the correlation between Cue 1 and the criterion (VIRUS LEVEL); $r_2$ is the correlation between Cue 2 and the criterion; $b_1$ and $b_2$ are the beta weights of Cue 1 and Cue 2, respectively; $R_e^2$ is the multiple correlation between the cues and the criterion.

---

[1] This research was supported by National Institute of Mental Health Grant MH-16437. The authors wish to express their appreciation to Berndt Brehmer and P. J. R. Boyle.

[2] Requests for reprints should be sent to Brian A. Knowles, Institute of Behavioral Science, University of Colorado, Boulder, Colorado 80302.

Since it was possible that the order of presentation of test trials would affect *Ss*' judgments of likely or unlikely, three different presentation orders of the test cases were used. Of the nine *Ss* in each of the four training conditions, three responded to each of the three presentation orders. For each presentation order, the orthogonality of cues was maintained for each block of 10 trials.

The *Ss*' judgments of the test trials as being likely or unlikely to occur provided the basic data for analysis. Specifically, four response measures were derived for each *S*: (*a*) the correlation between the cue values for those trials judged likely to occur; (*b*) the difference between this subjective cue intercorrelation and the actual cue intercorrelation in the task on which he was trained; (*c*) the standard error of estimate for predicting one cue from the other in those trials judged likely to occur; and (*d*) the ratio of this subjective standard error of estimate to the actual standard error in the training task. Each response measure was treated according to a 2 (Sign) × 2 (Magnitude) analysis of variance.

*Results.*—The analysis of the subjective intercorrelations transformed to Fisher's *Z* scores revealed only a significant main effect for sign of cue intercorrelation, $F\ (1,\ 32) = 96.59$, $p < .001$. As can be seen in Table 2, this effect indicates that *Ss* had learned a cue intercorrelation with the correct *sign*. That is, *Ss* in the positive conditions judged as "likely to occur" a set of instances in which the cues were positively correlated; *Ss* in the negative conditions selected a set of instances in which cues were negatively correlated. Indeed, 32 of 36 *Ss* estimated a correctly signed intercorrelation. Of the remaining 4 *Ss*, 3 estimated intercorrelations which were essentially zero. The absence of a significant main effect for magnitude of cue intercorrelation is due to the fact that optimal responding by *Ss* in all four conditions would result in a mean subjective intercorrelation of zero for *both* high redundancy (.8) and low redundancy (.4) conditions.

Analysis of the difference scores yielded only a significant effect for magnitude, $F\ (1,\ 32) = 9.76$, $p < .005$. That is, *Ss* in the .8 conditions departed more from optimal estimation than did *Ss* in the .4 conditions. Specifically, *Ss* in the .8 conditions tended to underestimate the magnitude of the cue intercorrelation (see Table 2). Although *Ss* in the positive intercorrelation conditions tended to be more accurate than *Ss* in the negative conditions, this sign effect did not attain a conventional level of significance, $F\ (1,\ 32) = 3.13$, $p < .10$.

TABLE 2

MEAN RESPONSE MEASURE ACCORDING TO EXPERIMENTAL CONDITION

| Cue intercorrelation in training | Response measure | | | |
|---|---|---|---|---|
| | $r_s$ | $r_t - r_s$ | $SE_s$ | $SE_s/SE_t$ |
| .8 | .69 | .16 | 2.74 | 1.18 |
| .4 | .55 | −.11 | 3.30 | .96 |
| −.8 | −.51 | .35 | 3.36 | 1.33 |
| −.4 | −.38 | .04 | 3.75 | 1.08 |

Note.—$r_s$ refers to the subjective cue intercorrelation; $r_t$ refers to the corresponding task intercorrelation; $SE_s$ indicates the subjective standard error of estimate; and $SE_t$ indicates the task standard error of estimate. The values of $r_t - r_s$ were corrected for sign so that a positive value always indicates underestimation of the *magnitude* of intercorrelation.

The analysis of the subjective standard error of estimate scores revealed significant main effects for both sign, $F\ (1,\ 32) = 10.42$, $p < .005$, and magnitude of cue intercorrelation, $F\ (1,\ 32) = 8.15$, $p < .01$. The sign effect resulted from the larger average standard error derived in the negative rather than in the positive redundancy conditions. The magnitude effect was due to the finding that *Ss* in the .4 conditions estimated larger standard errors than *Ss* in the .8 conditions. Both of these effects may be seen in Table 2.

Finally, the analysis of the ratio scores resulted in significant main effects for both sign, $F\ (1,\ 32) = 4.44$, $p < .05$ and magnitude, $F\ (1,\ 32) = 14.60$, $p < .001$. The sign effect was due to the larger ratios found in the negative conditions than in the positive conditions. The magnitude effect was due to the greater ratios found in the .8 conditions than in the .4 conditions. In short, *Ss* in the positive conditions were more accurate than those in the negative conditions, while *Ss* in the .4 conditions were more accurate than *Ss* in the .8 conditions (see Table 2).

*Discussion.*—From the results just described, two important findings emerge. First, it may be seen that *Ss* were able to detect an intercorrelation that they were not explicitly asked to learn. Second, not only were they able to detect the correct *sign* of the correlation, but they were reasonably accurate; their subjective standard errors of estimate were close to those that a statistician might specify for such correlations.

The accuracy of *Ss*' estimates, however, was influenced by both the magnitude and the sign of the cue intercorrelation. With regard to magnitude, it was found that although *Ss* in the high redundancy (.8) conditions correctly estimated a stronger relationship than *Ss* in the low redundancy (.4) conditions, *Ss* in the latter conditions were more accurate; specifically, *Ss* in the .8 conditions tended to underestimate the degree of relationship. With regard to sign, *Ss* in the positive conditions estimated a greater intercorrelation than *Ss* in the corresponding negative conditions; *Ss* in the negative conditions tended to underestimate the cue intercorrelation.

Although the present experiment concerned the (incidental) learning of a cue intercorrelation, the results are consistent with findings from studies in which *Ss* were instructed to learn the correlation between cue and criterion variables. Specifically, *Ss* (*a*) underestimate strong relations and overestimate weak relations and (*b*) learn positive relations more readily than negative ones (see Slovic & Lichtenstein, 1972, for a review). In short, the results of previous studies concerning the learning of cue-criterion relations in multiple cue probability tasks (e.g., Bjorkman, 1965; Peterson, Hammond, &

Summers, 1965) were found to generalize to the present study involving cue intercorrelations.

## REFERENCES

BJORKMAN, M. *Learning of linear functions; comparison between a positive and a negative slope.* (Rep. No. 183) Stockholm, Sweden: University of Stockholm, Psychological Laboratories, 1965.

KNOWLES, B. A., HAMMOND, K. R., STEWART, T. R., & SUMMERS, D. A. Positive and negative redundancy in multiple cue probability tasks. *Journal of Experimental Psychology* 1971, **90**, 157–159.

PETERSON, C. R., HAMMOND, K. R., & SUMMERS, D. A. Optimal responding in multiple cue probability learning. *Journal of Experimental Psychology*, 1965, **70**, 270–276.

SLOVIC, P., & LICHTENSTEIN, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. In L. Rappoport & D. Summers (Eds.), *Human judgment and social interaction.* Holt, Rinehart & Winston, 1972, in press.

(Received May 24, 1971)

---

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 427–429

# EFFECTS OF NOUN IMAGERY, PRONUNCIATION, METHOD OF PRESENTATION, AND INTRAPAIR ORDER OF ITEMS IN VERBAL DISCRIMINATION[1]

EDWARD J. ROWE[2] AND ALLAN PAIVIO

*University of Western Ontario*

Verbal discrimination learning (VDL) was investigated as a function of the imagery ($I$) value of items and three procedural variables. The *S*s received six trials on a 16-pair mixed list of High-$I$ and Low-$I$ words presented under (*a*) study-test or anticipation conditions, (*b*) pronunciation or no pronunciation of both pair members, and (*c*) constant or varied order of items within each pair. High-$I$ pairs were easier to discriminate than Low-$I$ pairs, and learning was facilitated when item pronunciation was not required. No other effects were significant, suggesting that method of presentation and intrapair order of items are less important than, and do not qualify the effects of, $I$ and pronunciation variables in VDL.

Paivio and Rowe (1970) found that verbal discrimination learning (VDL) was significantly easier for pairs of words rated high in imagery-arousing capacity ($I$) than for low-imagery words. By contrast, Thorndike-Lorge frequency (F) and associative meaningfulness ($m$) did not produce significant effects and, in a subsequent study (Rowe & Paivio, 1971), $I$ was more consistently effective than $F$ in both mixed and unmixed lists. In view of the apparent potency of the imagery attribute in VDL, it is important to determine the reliability of the effect under different experimental conditions. The above experiments were procedurally equivalent in all respects apart from the mixed-unmixed list variable, and it may be that the results obtained are specific to the particular procedures employed. An examination of the VDL literature suggests that two aspects of the experimental procedure are especially relevant in this respect: (*a*) a study-test procedure was used in all cases, and (*b*) *S*s were required to pronounce both the correct and incorrect items aloud on each study trial (where the correct item was underlined). The relevance of each of these factors is now considered.

Several investigators (e.g., Fulkerson & Johnson, 1971; Ingison & Ekstrand, 1970) have reported an improvement in VDL performance under study-test as opposed to anticipation presentation procedures. Ingison and Ekstrand also obtained no effect of concreteness on VDL, a result which has been attributed to the fact that their concrete and abstract words were not clearly differentiated on $I$ (see Rowe & Paivio, 1971). While improbable from an a priori point of view, it is possible that the imagery effect could be modified by differences in the method of presentation. The present experiment therefore included a comparison of High-$I$ and Low-$I$ word pairs under these two presentation conditions.

A general detrimental effect of item pronunciation in VDL has been consistently reported. Although a significant effect of $I$ has been found under both pronunciation and no-pronunciation conditions, degree of pronunciation and $I$ have not been systematically varied within the same experiment. Accordingly, this variable was included as a factor in the present experimental design.

A third factor, constant versus varied intrapair order of items, was also examined as a test of one possible explanation of the imagery effect. It is standard practice in VDL experiments to counterbalance the spatial position (left–right or top–bottom) of correct and incorrect items across acquisition trials. This procedure could serve to depress performance on Low-$I$ pairs, which are presumably processed predominantly via a verbal encoding system specialized for sequential processing of information. On the other hand, High-*I* words are encodable via an imaginal as well as a verbal system, and information storage and retrieval may not be affected to the same extent by sequential components of the task. Thus, the processing of Low-*I* items during VDL might be particularly sensitive to changes in the ordering of the members within each pair, whereas an alternative encoding system, i.e., visual imagery, is available for High-*I* pairs and these items should therefore be less affected by changes in intrapair ordering. Indirect evidence in support of this notion is available from paired-associate learning, where High-*I* pairs are less affected by recall order (forward or backward) than Low-*I* pairs.

---

[1] This research was supported by a National Research Council of Canada (NRC) Postgraduate Scholarship to the first author, and grants from National Research Council (APA-0087) and the University of Western Ontario Research Fund to the second author. We thank Ann Anas for her assistance in testing *S*s.

[2] Requests for reprints should be sent to Edward J. Rowe, who is now at the Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

*Method.*—Thirty-two, one- and two-syllable words, four to seven letters in length, were selected from the Paivio, Yuille, and Madigan (1968) norms. Half of the words (High-*I*) were rated above 6.00 on $I$ ($\bar{X} = 6.35$) and the remainder (Low-*I*) had ratings of less than 3.68 ($\bar{X} = 3.18$). All items were greater than 18 per million by the Thorndike-Lorge count and the High-*I* and Low-*I* words were approximately matched on $m$. Eight pairs of each imagery type were constructed by a nonsystematic procedure and the two sets combined to form a 16-pair mixed list (List 1). A second list of 16 different pairs (List 2) was obtained by an identical procedure.

The *S*s were 96 introductory psychology students (55 males) who participated in the experiment as part of a course requirement. None had prior experience with the VDL task. The design was a $2^4$ factorial with pronunciation, method (anticipation and study-test) and order (constant and varied) as between-*S*s factors and imagery as a within-*S*s factor. Twelve *S*s were assigned in rotation to each of the eight independent groups.

Each *S* first received conventional VDL instructions and one trial on a four-pair practice list of Medium-*I* words. The practice pairs were typed on index cards and presented manually by *E* using either a study-test or anticipation procedure to match the testing conditions of the remainder of the experiment. The experimental lists were presented on a Gerbrands memory drum at a 3-sec. rate with a 3-sec. intertrial interval. For the study (and feedback) presentations, one randomly selected item was underlined. On the test (and anticipation) presentations, the underlining was absent, and *S* chose what he thought was the correct word for each pair. All *S*s were instructed to guess when uncertain. Learning proceeded for a total of six trials. The *S*s in the pronunciation conditions pronounced both words aloud, reading from left to right, on the study and feedback presentations, while *S*s in the no-pronunciation condition were silent. With the anticipation procedure, the first trial was preceded by a trial in which all 16 pairs were presented with the correct items underlined. Since the first presentation in the study-test condition was also a study trial, this insured that all *S*s received the same initial exposure to the list. For the constant order condition, the pairs of words occupied the same spatial position (left or right) across all presentations of the list, whereas for the varied order condition, half of the pairs were reversed on each successive presentation. There were four different random orders of the pairs in the study-test condition, and it was possible to carry out the reversal procedure so that the members of any particular pair appeared equally often in the left and right positions across all six trials. With the anticipation procedure, only three different orders were possible because of mechanical restrictions involved in the construction of the memory drum tapes, and left–right position could therefore not be completely counterbalanced within a given pair. Nevertheless, the items of each pair changed position three times over trials in the varied order condition, so the intrapair position of items could not be used consistently as an aid in learning.

To increase the generality of the results, half of the *S*s in each condition received List 1 and the other half, List 2. Furthermore, for half of the *S*s receiving each list, one item of each pair was correct while for the remaining half the other item was correct.

*Results and discussion.*—The mean total number of errors over trials for each condition is given in Table 1. A four-way analysis of variance resulted in significant main effects for both imagery, $F(1, 88) = 44.48$, $p < .001$, and pronunciation, $F(1, 88) = 17.29$, $p < .001$. High-*I* pairs produced fewer errors than Low-*I* pairs (5.81 vs. 9.49), and fewer errors occurred in the no-pronunciation condition (5.47 vs 9.83). The pronunciation factor accounted for about 1.36 times the variance accounted for by imagery. Errors were fewer with the study-test than with the anticipation procedure (6.97 vs. 8.33), and for constant as opposed to varied intrapair order (6.80 vs. 8.50), but these differences did not approach significance. Also, none of the interactions were significant.

TABLE 1

MEAN NUMBER OF ERRORS FOR EACH EXPERIMENTAL CONDITION

| Pair type | Presentation cond. | | | |
|---|---|---|---|---|
| | Anticipation | | Study test | |
| | Constant order | Varied order | Constant order | Varied order |
| | Pronunciation | | | |
| High-*I* | 8.42 | 7.42 | 6.25 | 8.67 |
| Low-*I* | 14.17 | 10.08 | 10.33 | 13.33 |
| | No pronunciation | | | |
| High-*I* | 2.92 | 6.25 | 2.33 | 4.25 |
| Low-*I* | 5.67 | 11.75 | 4.33 | 6.25 |

Note.—The standard deviations ranged from 3.27 to 7.03.

The conclusions are straightforward. The fact that the difference between High-*I* and Low-*I* pairs was not significantly affected by changes in presentation procedure or pronunciation conditions is further evidence that the effect of *I* in VDL is highly reliable. The present results similarly concur with previous findings in demonstrating a strong effect of pronunciation. By contrast, study-test versus anticipation methods of presentation and constant versus varied intrapair ordering of items appear to be relatively unimportant factors in the task.

Since a significant difference favoring study-test over anticipation methods has been found by previous investigators (e.g., Fulkerson & Johnson, 1971; Ingison & Ekstrand, 1970), these studies were examined for procedural variations which might account for our disparate results. The experiments differ in a number of ways, including type of items, list length, presentation rate, method of feedback presentation, and the presence or absence of an initial study trial, but the superiority of the study-test procedure seems unaffected by these factors. One way in which the present study departs from the others is in the use of a fixed number of trials (six) rather than a criterion requirement. At the end of the six trials, only 10 *S*s in the anticipation condition and 17 in the study-test condition had attained a criterion of two consecutive errorless trials, thus raising the possibility that a significant difference between the two procedures might have been observed had performance been taken to this criterion for all *S*s. This possibility cannot be tested directly by the present results, but two observations argue against it. First, if the interpretation is correct, scores for the study-test group would be expected to show a tendency to attain criterion faster than anticipation scores across trials. An examination of the learning curves for the two groups, with the data collapsed across the other conditions, showed no evidence to support this expectation. If anything, the curve for the anticipation group was steeper, but an analysis of variance showed that the Groups × Trials interaction was not significant. Second, Fulkerson and Johnson (1971) have reported a significant effect of method of presentation even when only the error data for early trials were considered, suggesting that their findings, at least, were not due to the criterion level of performance attained by their *S*s. Our failure to demonstrate a facilitation under study-test conditions thus remains unexplained, but the results of the present experiment do suggest that differences in method of presentation are less important in VDL than variation in imagery and pronunciation.

Finally, the hypothesis that learning of Low-*I* pairs in VDL is adversely affected by varying the intrapair order of items was not confirmed. Changing the intrapair order from trial to trial did not significantly affect performance. It would appear, therefore, that the two types of items are processed independently of their left–right spatial position, at least when position of the correct item is randomized between pairs. The disconfirmation of the hypothesis suggests that a satisfactory explanation of the effect of *I* is most likely to be found in the different characteristics of High-*I* and Low-*I* words themselves. Evidence supporting an interpretation of the effect in terms of the differential use of imagery for the two word types has been presented elsewhere (e.g., Rowe & Paivio, 1971). Such evidence encourages further research designed to test more specific hypotheses concerning the role of imagery in VDL involving High-*I* pairs.

## REFERENCES

FULKERSON, F. E., & JOHNSON, J. E. Methodological variables in verbal discrimination learning. *Psychonomic Science*, 1971, **22**, 68–69.

INGISON, L. J., & EKSTRAND, B. R. Effects of study time, method of presentation, word frequency, and word abstractness on verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **85**, 249–254.

PAIVIO, A., & ROWE, E. J. Noun imagery, frequency, and meaningfulness in verbal discrimination. *Journal of Experimental Psychology*, 1970, **85**, 264–269.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76**(1, Pt. 2).

ROWE, E. J., & PAIVIO, A. Word frequency and imagery effects in verbal discrimination learning. *Journal of Experimental Psychology*, 1971, **88**, 319–326.

(Received June 26, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 430–432

# SEMANTIC GENERALIZATION: REPLICATION OF CRAMER'S ASSOCIATIVE GRADIENT[1]

S. JOYCE BROTSKY[2] AND LINDA M. LITWIN

*San Fernando Valley State College*

Fifty-two *S*s were shown 24 different Critical Stimulus (CS) words. Following training, *S* was shown the 24 original words, 96 associatively related generalization stimulus (GS) words, and 24 unrelated words and was asked to press one of six buttons to indicate his degree of confidence that he had seen each of the words during training. Normative associative strength between CS and GS predicted a monotonic gradient of semantic generalization when a button-press confidence rating was used as the dependent variable.

If a response is conditioned to a word (critical stimulus), that response should generalize to semantically related words. Moreover, the amount of generalization should vary with the strength of the relationship between the critical stimulus (CS) and the generalization stimulus (GS). Although semantic generalization has been demonstrated using either an autonomic response or a voluntary response as the dependent variable (Brotsky, 1968; Maltzman & Raskin, 1965; Mednick, 1957; Mink, 1963), successful demonstrations of an associative gradient predicted from normative free association strength (FAS) have been sparse (Kurcz, 1964; Maltzman, 1968). Cramer (1970), however, has recently demonstrated a monotonic associative gradient using overt and partial muscle responses to the CS and the GS as her dependent variable. The present study was an attempt to replicate and expand on Cramer's results using a simple but highly sensitive response, a button-press confidence rating, as the dependent variable.

*Method.*—The *S*s were 26 men and 26 women enrolled in introductory psychology courses at San Fernando Valley State College whose participation fulfilled a course requirement. All *S*s were between the ages of 17 and 27, were born in continental United States, and spoke English as their first language. Data from 8 additional *S*s were discarded. There were two *E*s ($E_1$ and $E_2$) each of whom ran 13 men and 13 women *S*s.

All stimulus words were taken from Cramer's (1970) study. There were 24 CS words, and for each of the CS words there were 4 related GSs of decreasing normative FAS and an unrelated control word (X) constituting a list of 144 stimulus words. The training list consisted of 24 CSs presented in a different random order to each *S*, with 1 buffer word at the beginning and end of the training series. For the test series that followed, the words were arranged in blocks of 6, 1 word from each category (CS, $GS_1$, $GS_2$, $GS_3$, $GS_4$, and X), randomly selected and ordered following the arrangement used by Cramer. The words were printed on transparent plastic, inserted in 2 × 2 in. plastic slide mounts, and projected with an overall slide image of 6 × 9 in., the letters appearing $\frac{3}{4}$ in. high.

The *S*s were seated in a 4 × 7 × 7 ft. plywood cubicle with a table and a Polacoat 18 × 18 in. screen before them. The slides were projected from outside the cubicle (rear-projection method) at eye level via a Kodak Carousel AV-900 projector. On the table was a metal box consisting of six push-buttons marked 1, 2, 3, 4, 5, and 6 across the upper surface. Instructions, prerecorded by $E_1$ (L. L.) were presented via earphones that were worn by *S* throughout the experiment. A continuous low-frequency white noise (about 20 db.) provided auditory masking at all times. During the training, *S*s were asked to press the key marked 6 immediately at the presentation of each of the CS words. The *S*s were not explicitly instructed to learn the words, since the visual stimuli were ostensibly presented to test their reaction time. All slides were presented at a rate of 2-sec. exposure per slide automatically timed. Following presentation of the training words, a series of 65 two- and three-digit numbers was presented auditorially via the earphones at a 5-sec. rate, and *S* was asked to subtract the digit 3 from each number, writing the answers on a numbered sheet provided him. The subtraction task filled a 5-min. period between the training series and the presentation of the test list.

For the test list, *S*s were asked to press a key as quickly as possible in response to the presentation of each word, using the keys to make a confidence rating as to whether the word had appeared on the first list. Maximum confidence that the word had appeared in the first series was expressed with Key 6, maximum confidence that it had not previously appeared was expressed with Key 1. Intermediate degrees of confidence were to be reflected by the use of the middle keys. Speed of response was stressed during the course of the instructions, and if *S* made his first or second response to the words in the test list after the presentation of the next slide in the series, he was instructed "please respond more quickly." The data from five *S*s were discarded because *S*s did not respond prior to the presentation of the next slide on seven or more occasions.

A time interval generator was used in connection with a two-channel Offner dynagraph, wired to the projector and to the response mechanism. One channel was used to record the mechanical presentation of the slide, while the second channel was used to record *S*'s response. The pen was set to move 1 unit horizontally to the left of the base line or 5 mm. for a response with Key 1, 10 mm. for a response with Key 2, etc. The response keys were automatically reset when the slide changed and the polygraph pen automatically returned to the base line.

---

[1] This work was supported by a Faculty Research Grant to the senior author from the San Fernando Valley State College Foundation. The authors are indebted to Donald C. Butler for assistance in computer programming and to Leslie Shaps for assistance in data collection.

[2] Requests for reprints should be sent to S. Joyce Brotsky, San Fernando Valley State College, Northridge, California 91324.

*Results.*—The mean confidence ratings during the test to the 24 words in each of the six stimulus categories (CS, GS$_1$, GS$_2$, GS$_3$, GS$_4$, and X) were determined for each *S*. The mean confidence ratings of the four groups of *S*s are presented in Fig. 1. A $2 \times 2 \times 6$ analysis of variance was conducted where *E* and sex of *S* were the between-groups factors and word category was the within-groups factor. Neither the main effect of sex nor *E* was statistically reliable, but the Sex $\times$ *E* interaction was statistically significant, $F\,(1, 48) = 4.80$, $p < .05$. That is, although the women's overall level of confidence ratings was consistent regardless of *E*, the men showed a lower mean level of confidence rating with $E_1$ (a woman) but a higher mean level of confidence rating with $E_2$ (a man). Word category was statistically significant, $F\,(5, 240) = 114.09$, $p < .01$.

Using the within-groups error mean square from the analysis of variance (.154), Dunnett's *t* tests were performed to provide a comparison of the individual word category means with the control mean. The word category means collapsed over sex of *S* and *E* are presented in Table 1. Dunnett's *t* indicated that there was a significant difference between the magnitude of the confidence rating to X and to CS words, $t\,(240) = 20.78$; to GS$_1$, $t\,(240) = 9.09$; to GS$_2$, $t\,(240) = 6.49$; to GS$_3$, $t\,(240) = 3.90$; and to GS$_4$, $t\,(240) = 2.59$; all $p\text{s} < .05$.

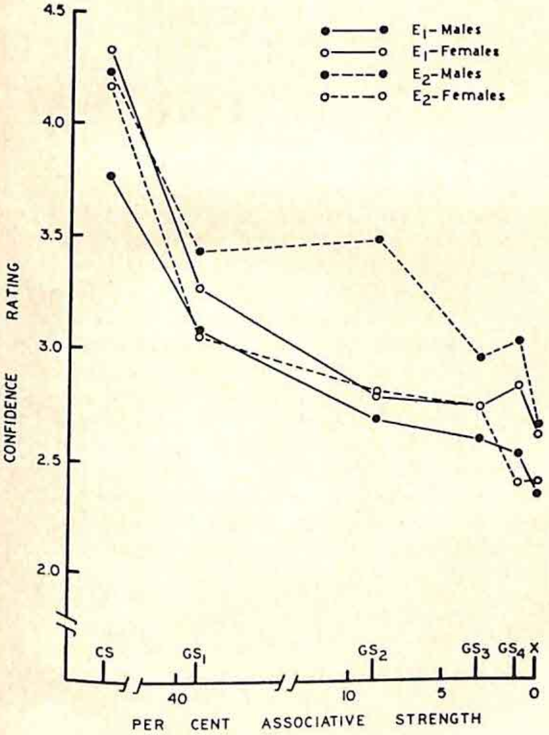


FIG. 1. Magnitude of the button-press confidence ratings to the CS, the GSs, and the X word in the test for the men and women run by the two *E*s.

TABLE 1

MEAN MAGNITUDE OF THE BUTTON-PRESS CONFIDENCE RATING TO THE SIX CATEGORIES OF WORDS FOR THE FIFTY-TWO SUBJECTS

| Word category | | | | | |
|---|---|---|---|---|---|
| CS | GS$_1$ | GS$_2$ | GS$_3$ | GS$_4$ | X |
| 4.125 | 3.203 | 2.943 | 2.760 | 2.699 | 2.508 |

Note.—CS = critical stimulus, GS = generalization stimulus, and X = an unrelated control word.

These results indicated that incidental learning did take place (CS > X) and that semantic generalization had occurred.

Since a great part of the word category variance reflected the confidence ratings to CS words, the scores based on these responses were removed and the data were reanalyzed for the responses to generalization and control stimuli only. A $2 \times 2 \times 5$ analysis of variance indicated that neither sex of *S* nor *E* was statistically reliable, but the Sex $\times$ *E* interaction was statistically significant, $F\,(1, 48) = 4.53$, $p < .05$, as in the previous analysis. The main effect for word categories was statistically significant as predicted, $F\,(4, 192) = 28.62$, $p < .01$.

To demonstrate a gradient of semantic generalization, it was necessary to determine if there were differences in rating responses to words in adjacent stimulus categories. Using the within-groups error mean square from the previous analysis (.126), a Duncan's multiple-range test indicated that every category mean differed significantly from every other category mean, ($p < .05$). An orthogonal polynomial trend analysis was conducted on the confidence ratings to the four GSs to determine the approximate shape of the generalization gradient.[3] The significant linear component of the mean, $F_{\text{lin}}\,(1, 48) = 52.53$, $p < .01$, indicated that generalization was a linear function of the associative strength between the CS and the associated GSs.

*Discussion.*—Over the last decade, many studies have investigated semantic generalization using the classical conditioning paradigm or the instrumental conditioning paradigm. Although more test list generalization has typically occurred to related words than to unrelated words, the predicted relationship between FAS and magnitude of generalization has been demonstrated only by Cramer (1970). Two factors may account for the earlier failures to demonstrate the associative gradient: (*a*) Both autonomic responses or confidence ratings where *S* responds YES or NO are inadequate dependent variables because they are not sufficiently sensitive, and (*b*) FAS alone is not an adequate predictor of semantic generalization because additional stimulus characteristics must be considered.

---

[3] Percent FAS was not used in the trend analysis. Instead, an equal interval trend analysis was arbitrarily carried out to ascertain the approximate form of the generalization gradient.

It seems probable that the YES or NO response is relatively insensitive in comparison with a dependent variable based on the exact confidence ratings produced by *S*. Although the assumption can be made that there is an underlying distribution of confidence ratings where YES and NO represent the extremes, it seems more appropriate to use a distribution which is generated from *S*'s data using several points than to use a distribution which is generated from *S*'s data using only two points. Moreover, the present study tended to promote semantic generalization through the use of incidental learning instructions and a fast rate of stimulus presentation (Cramer, 1971).

The appropriateness of FAS as a predictor of semantic generalization must also be questioned. Although semantic generalization and paired-associate learning (PAL) should vary with normative FAS, the data have not conformed to prediction for either task. Linton, Jewett, and Brotsky (1970) were able to predict PAL from FAS when they used a complex indicator, FAS and normative rank. Cramer (1970) recognized the complexity of the problem, for although the FAS of the words in her stimulus pool was cited, additional criteria were cited, the most salient of which was probably normative rank. In all cases, $GS_1$ was the normative primary response to CS, $GS_2$ was the secondary normative response to CS, and rank was considered for $GS_3$ and $GS_4$ as well. Further research must be concerned with the demonstration of associative gradients using other samples of stimulus words.

## REFERENCES

BROTSKY, S. J. Classical conditioning of the galvanic skin response to verbal concepts. *Journal of Experimental Psychology*, 1968, **76**, 244–253.

CRAMER, P. Semantic generalization: Demonstration of an associative gradient. *Journal of Experimental Psychology*, 1970, **84**, 164–172.

CRAMER, P. Discrimination as a factor in semantic generalization. *Journal of Experimental Psychology*, 1971, **87**, 396–400.

KURCZ, I. Semantic and phonetographic generalization of a voluntary response. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 261–268.

LINTON, M., JEWETT, M., & BROTSKY, S. J. Prediction of paired-associate learning from a complex indicator of associative strength. *Psychonomic Science*, 1970, **21**, 354–355.

MALTZMAN, I. Theoretical conceptions of semantic conditioning and generalization. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice Hall, 1968.

MALTZMAN, I., & RASKIN, D. C. Effects of individual differences in the orienting reflex. *Journal of Experimental Research in Personality*, 1965, **1**, 1–16.

MEDNICK, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1957, **55**, 315–321.

MINK, W. D. Semantic generalization as related to word association. *Psychological Reports*, 1963, **12**, 59–67.

(Received July 29, 1971)

*Journal of Experimental Psychology*
1972, Vol. 93, No. 2, 433–434

# STIMULUS ANTICIPATION IN CHOICE REACTION TIME WITH VARIABLE S—R MAPPING[1]

CHARLES P. WHITMAN AND E. SCOTT GELLER[2]

*Virginia Polytechnic Institute and State University*

The effect of prediction outcome on choice reaction time was studied when each stimulus presentation consisted of a rule bit which defined the S–R mapping and a data bit which determined the response. Prior to each presentation, *S*s predicted both the data and rule portions of the signal. Choice latencies were shortest when both the data and rule were correctly predicted and longest when both were incorrectly predicted. The results indicated that *S*s do not anticipate the response implied by the predicted stimulus.

Recent choice reaction time (RT) studies have employed the three-stimulus, two-response design introduced by LaBerge and Tweedy (1964) to determine the relative importance of stimulus anticipation and response preparation when *S*s' react faster to correctly predicted events (e.g., Hinrichs & Krainz, 1970; Whitman & Geller, 1972). The general conclusion was that the prediction outcome (PO) effect was primarily due to stimulus anticipation. Although these experiments offered interesting results, they failed to completely address the question. That is, their design could study choice RT following three of the four possible combinations of stimulus and response PO. The *S*s could correctly predict the stimulus and the response, *S*s could incorrectly predict the stimulus while correctly anticipating the response, or *S*s could incorrectly predict both the stimulus and the response. Thus, an inadequacy of the above design was that *S*s could not correctly predict the stimulus while incorrectly anticipating the response.

Perhaps a solution to this problem could be found in the utilization of a design introduced by Shaffer (1965) and later used by LeMay and Simon (1969) in which *S*s responded to a one-bit "data" stimulus according to the S–R mapping indicated by one-bit "rule" stimulus. With the addition that *S*s predict both the rule and the data of each stimulus presentation, the present experiment determined the effects of all four possible PO combinations on choice RT.

*Method.*—The 10 *S*s were volunteers from psychology classes at Virginia Polytechnic Institute and State University. All *S*s had previously participated in a two-choice RT experiment with one-to-one mapping of stimulus to response.

The two data alternatives were the symbols $U$ and $\Omega$, verbally labeled "up" and "down." The two rule alternatives consisted of a dot which appeared at the top or bottom (labeled "top" and "bottom") of the same 1.3 × 2 cm. digital readout which presented the data. The four possible combinations of the data and rule were equiprobable and the random sequence was determined by an IBM 360 computer system. The *S*s reacted by pressing one of two telegraph keys: for five *S*s a top dot indicated a right response to $U$ and a left response to $\Omega$, while a bottom dot indicated a left response to $U$ and a right response to $\Omega$; the remaining five *S*s reacted with the opposite relationship between rule and response mapping. Incorrect identifications were reported to *E* by indicator lights and these RTs were deleted from the analysis. Choice RTs were measured to the nearest millisecond by a Hunter Klockcounter.

Each *S* reacted to 150 presentations during one session and to 200 presentations during a second session, for a total of 350 trials. The sequence of events on each trial was as follows: *S*'s verbal prediction, a brief "ready" buzzer, a random time interval of between .5 and 1 sec., a simultaneous presentation of data and rule, and *S*'s identification response stopped the clock and turned off the stimulus presentation.

*Results and discussion.*—The 4 PO (prediction outcome) categories presented in Fig. 1 were as follows: a correct prediction of both the data and rule, a correct prediction of only the rule, a correct prediction of only the data, or a correct prediction of neither data nor rule. For each *S*, the mean latency of each category was calculated and used in the analyses presented below.

As indicated in Fig. 1, a correct prediction of both bits of information resulted in a shorter latency than a correct prediction of either bit alone. An analysis of variance, 2 (PO of the Data) × 2 (PO of the Rule), indicated significant main effects of rule PO, $F(1, 9) = 7.51$, $p < .05$, and data PO, $F(1, 9) = 6.91$, $p < .05$. Figure 1 also indicated that correctly anticipating either bit had a greater effect on choice RT when the other bit was correctly predicted than when the other bit was incorrectly predicted, and so the Rule PO × Data PO interaction was reliable, $F(1, 9) = 8.32$, $p < .05$. Although the figure suggests shorter latencies for correct rule POs than for correct data POs, the difference was not significant ($t = 1.05$, $p > .10$, two-tailed).

It should be noted that a correct prediction of the S–R rule was not always equivalent to a correct prediction of the response. When the data was correctly predicted the response PO was indeed the rule PO, but when the data was incorrectly pre-

---

[1] This research was supported in part by the Small Grants program of the College of Arts and Sciences of Virginia Polytechnic Institute and State University. The authors are indebted to John C. Farris and David S. Post for their aid in collecting the data.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

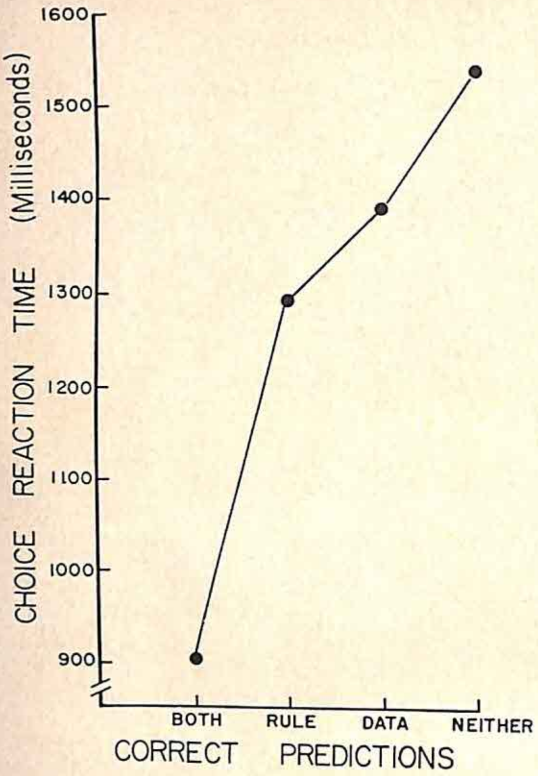


FIG. 1. Choice reaction time as a function of the prediction outcome of the rule and data bits of the stimulus presentation.

dicted the response PO was the opposite of the rule PO. More specifically, a correct prediction of both the data and the rule resulted in a correct prediction of the response, and a correct prediction of only the data resulted in an incorrect prediction of the response; but a correct prediction of only the rule resulted in an incorrect prediction of the response, and an incorrect prediction of both the data and the rule resulted in a correct prediction of the response. Since choice RT was an increasing function of the number of components of the stimulus configuration which were incorrectly predicted, choice RT was longest when both the data and the rule were incorrectly predicted even though in this case the actual response was correctly anticipated.

Previous investigators (Hinrichs & Krainz, 1970; Whitman & Geller, 1972) have examined the notion that response anticipation could account for PO effects. This hypothesis was founded on the possibility that *S*s process their stimulus predictions and subsequently prepare to make the anticipated response prior to the stimulus presentation. By indicating that *S*s do not anticipate or prepare for the response implicated by their stimulus prediction, the results of the present study support the conclusions of these other investigators.

## REFERENCES

HINRICHS, J. V., & KRAINZ, P. L. Expectancy in choice reaction time: Anticipation of stimulus or response? *Journal of Experimental Psychology*, 1970, **85**, 330–334.

LABERGE, D., & TWEEDY, J. R. Presentation probability and choice time. *Journal of Experimental Psychology*, 1964, **68**, 477–481.

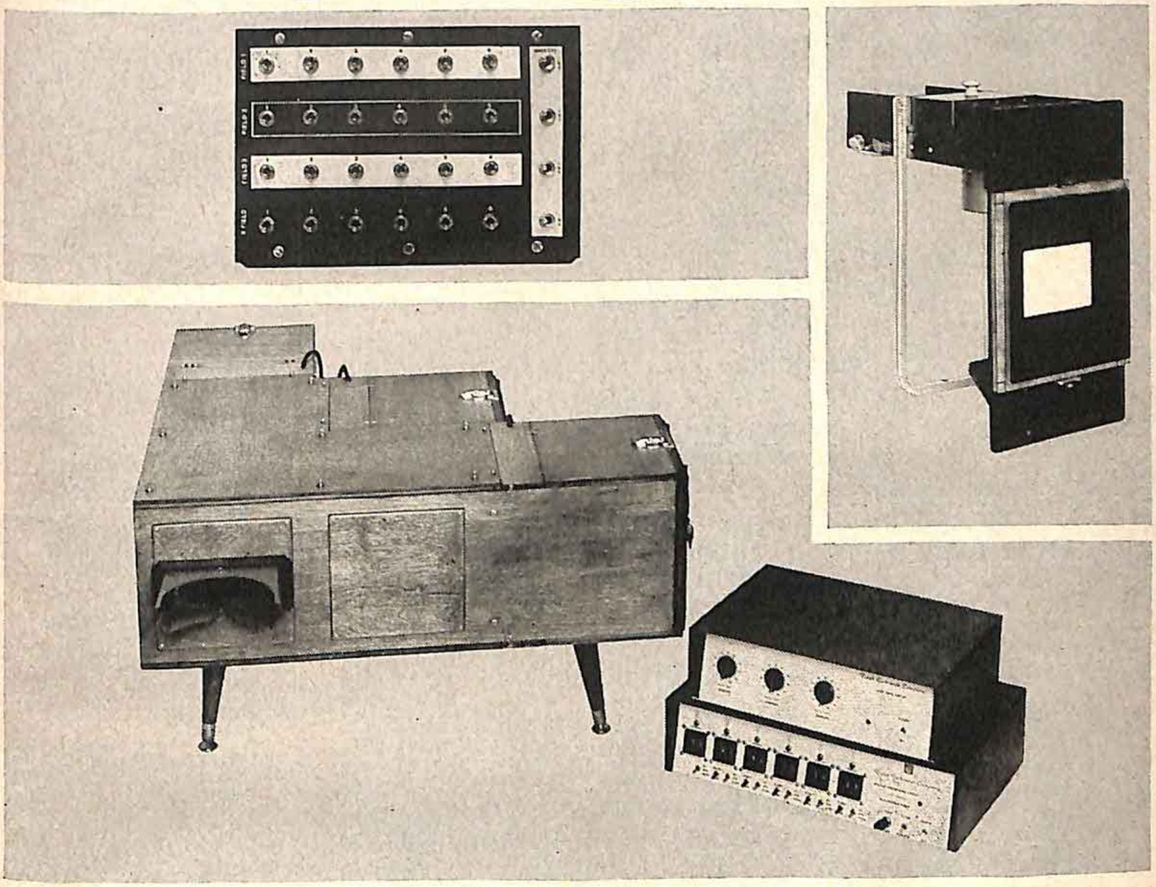
LEMAY, R. P., & SIMON, J. R. Temporal and symbolic S–R compatibility in a sequential information-processing task. *Journal of Experimental Psychology*, 1969, **80**, 558–560.

SHAFFER, L. H. Choice reaction with variable S–R mapping. *Journal of Experimental Psychology*, 1965, **70**, 284–288.

WHITMAN, C. P., & GELLER, E. S. Prediction outcome and choice reaction time: stimulus versus response anticipation. *Journal of Experimental Psychology*, 1972, **93**, 193–197 .

(Received October 26, 1971)

# at last ! easy alignment



That's right! For ease of alignment, all our tachistoscopes now have a new *center pivoted, gimbal mount* mirror system.

What else? Well, we have a new Universal Tachistoscope Logic Interface which enables you to program complex tachistoscopic functions. We also have a new Automatic Card Changer which attaches directly to the tachistoscope cabinet.

And, don't forget. All our Harvard Tachistoscopes come complete with direct reading electronic digital timers [accuracy ±.005%], solid-state lamp drivers, and interchangeable stimulus backs.

Versatile, reliable, and easy to align. Write for complete information.

**Ralph Gerbrands Company**
8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

# ANNOUNCING

# OUR NEW INSTANT LABORATORY

It's called the Behavior Observation Kit. It was designed for Rinehart Press by Professors Robert L. Marrone and Richard Rasor and constructed by Macalaster Scientific Company of high-grade transparent injection mold plastic to guarantee durability.

With this completely portable and inexpensive laboratory, students will be able to conduct behavior experiments anywhere. Behavior Observation Kits can turn any classroom into a low-cost laboratory and students can take kits home to conduct their own experiments.

The Kit consists of seven components: an operant chamber (automated and equipped for studies in reinforcement, extinction, discrimination learning, avoidance learning, escape conditioning, perception, motivation); a partitioned, gridded observation chamber and aquarium; a small animal visual cliff apparatus; a worm or ant farm; a mirror tracing apparatus; an insect-worm maze; and a human punch maze and pencil maze. This flexibility enables students to perform a wide variety of animal and human experiments.

The versatile, durable, portable
**BEHAVIOR OBSERVATION KIT**
*Designed by* Robert L. Marrone, *Sacramento State College, and* Richard Rasor, *American River College*
$29.95 (f.o.b.)

Interested in starting your own instant laboratory? Write for more information to: Dept. EP,

**RINEHART PRESS**
5643 Paradise Drive
Corte Madera, California 94925

# The Projection T-Scope

## ...how we can make it easier for you to instrument and control your experiments.

An ICONIX projection T-scope system can be made from the following modules: A Random Access Projector with Controller, to present any order of slide stimuli to your subjects; Electronic Shutter and Shutter Driver, for precise and reliable control of stimuli; Timer and Controller, to set a stimulus presentation order and time intervals, and a Paper Tape Reader with Controller, to provide a repeatable presentation of a pre-programmed stimulus sequence.

Here are some features of these products:

**Random Access Projector with Controller:**
Model 6189—$1350.

1. 80 Slide Kodak Projector (RA 960) has control unit and mode switching in rack mounted cabinet.
2. Front panel operation in manual mode, automatic mode controlled by buffered tape reader.
3. Maximum slide search cycle—3.5 seconds.

Note: We also have serial projectors from $105.

**hutters:** Model 1425, 25mm Aperture—$85.
Model 1427, 6mm Aperture—$85.

1. 6mm aperture shutter allows a minimum effective exposure time of 1.5 milliseconds.
2. 25mm aperture shutter passes four f-stops more light than the 6mm model, with a minimum effective exposure time of 5 milliseconds.
3. Shutters built for extreme reliability, with a useful life of over half a million cycles.

**Shutter Drivers:** Model 6188-2, two channel—$425.
Model 6188-3, three channel—$500.

1. Operate electronic shutters from either a pulse or a DC level input.
2. Repetition rate of >2 presentations per second.
3. Included power supplies can provide power to other equipment.

**Punched Tape Reader with Controller:**
Model 6319—$775.

1. Reads punched tape program at a 50 cps rate, and transfers standard 4-bit codes to optional 16-, 32-, or 64-bit program buffers.
2. Reads one 8-bit non-standard paper tape code at rest, effectively and 8-bit buffer.

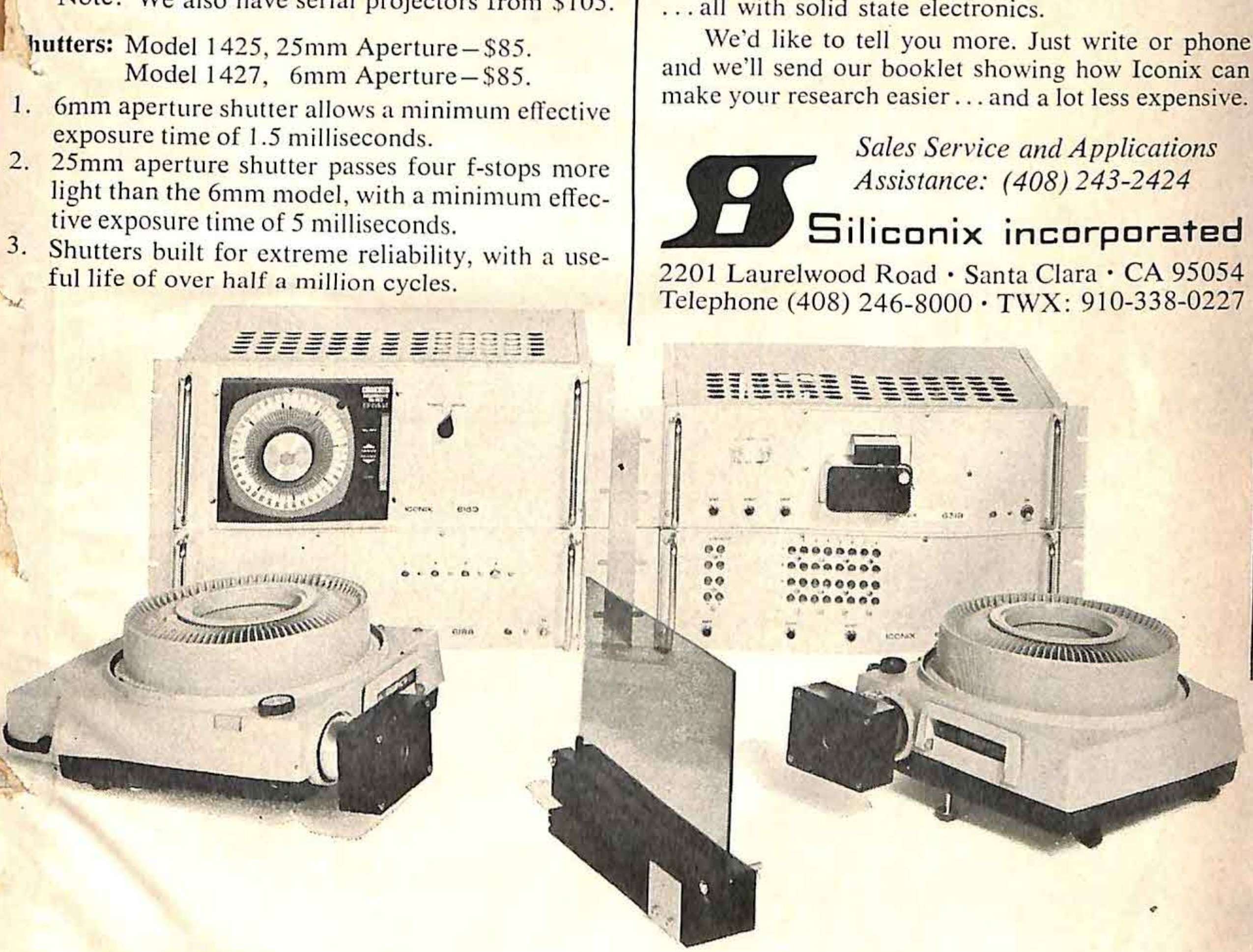
You should know more about *Iconix Products*. We make *versatile timer/control units*, a highly flexible *logic cabinet with program plugboard*, and a standard line of over 30 proven and reliable products ...all with solid state electronics.

We'd like to tell you more. Just write or phone and we'll send our booklet showing how Iconix can make your research easier... and a lot less expensive.

*Sales Service and Applications Assistance: (408) 243-2424*

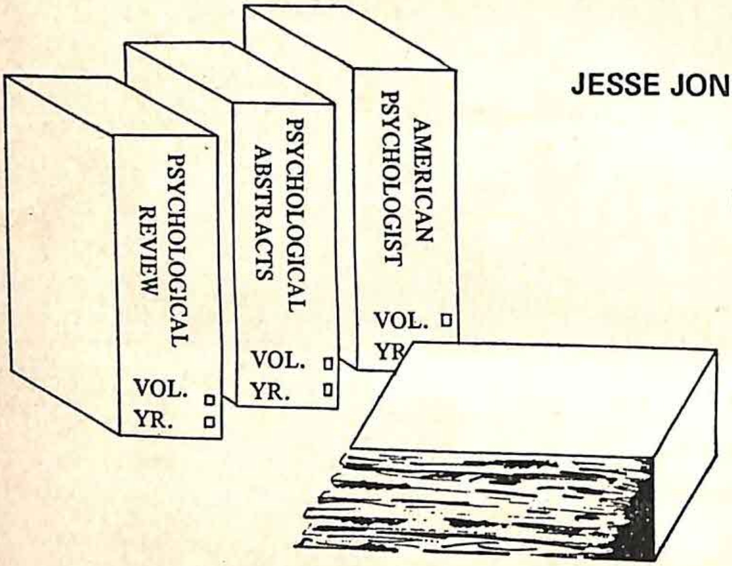
**Siliconix incorporated**

2201 Laurelwood Road · Santa Clara · CA 95054
Telephone (408) 246-8000 · TWX: 910-338-0227



JESSE JONES VOLUME FILES

Send orders to:
JESSE JONES CORP.
P.O. Box 5120
Philadelphia, Pa. 19141



Only $3.50 ea.
3 for $10.00
6 for $19.00 P.P.D.

Now, the journals of the American Psychological Association can become an attractive permanent part of your professional library. These famous Jesse Jones volume files, especially designed to keep your copies orderly, readily accessible for future references — guard against soiling, tearing, wear or misplacement of copies.

This durable file will support 150 lbs. Looks and feels like leather and is washable. The 23-carat gold lettering makes it a fit companion for the most costly binding.